

The meaning of life

'Pro-life' groups define the beginning of human life as the union of sperm and egg, and equate the harvesting of human embryonic stem cells to homicide. But our biological understanding lends little support to these views.

Last week, a private clinic in Virginia further inflamed the debate over embryonic stem (ES) cell research. Researchers at the Jones Institute for Reproductive Medicine in Norfolk announced that they had created human embryos by *in vitro* fertilization for the sole purpose of harvesting ES cells. 'Pro-life' groups denounced the clinic. "It's still killing a human being," a representative of the Virginia Society for Human Life told *The Virginian-Pilot* newspaper.

Central to this argument is the view that human life begins at fertilization. ES cells show promise in the field of regenerative medicine, which seeks to grow tissues to replace those lost to disease or injury. But they are harvested by destroying an embryo comprising a hollow ball of cells called a blastocyst. The idea that this represents the destruction of a human life has an appealing moral simplicity.

Biology, however, is not that simple. Recent advances in reproductive medicine have emphasized that mammalian life need not start with the union of sperm and egg. Fertilization is not required to create embryos by nuclear-transfer cloning, the technique used to produce Dolly the sheep.

And earlier this month at a meeting in Lausanne, Switzerland, an Australian team described experiments in which mouse eggs were 'fertilized' with cells taken from adult mice. The resulting embryos contained an extra set of chromosomes, but could be induced to expel them and begin to develop as normal.

Those who oppose the extraction of ES cells from human blastocysts also tend to oppose such experimental manipulations of reproductive biology. But the natural phenomenon of identical twins similarly creates problems for the simple view that human life starts at fertilization. An embryo can split to form two or more viable

embryos at any stage up to 'gastrulation', when its cells begin to migrate into distinct layers that form the basis of the adult body plan.

Because of twinning, many bioethicists take the view that it is only after gastrulation that an embryo can begin to be considered as an individual human being. This definition can be applied whether an embryo is created by conventional fertilization or by any other procedure. It does not deny that pre-gastrulation embryos are alive — but then so too, in some senses, are sperm and egg cells.

Framing the discussion over ES-cell research in the context of when biological individuality arises, rather than when life begins, could lead to a more meaningful debate. That is why it is disappointing to see one of the companies working on ES cells apparently buying into the 'life begins at fertilization' argument.

Advanced Cell Technology (ACT) of Worcester, Massachusetts, says it is trying to generate human embryos by cloning, and then harvest ES cells from them. The company hopes to sidestep moral objections, as fertilization is not involved. Indeed, the chair of ACT's ethical advisory board argues that an embryo created in this way is not a bona fide embryo, and suggests the term 'ovumsum'.

The procedure that ACT is experimenting with, known as therapeutic cloning, might one day prove useful in generating ES cells that are genetically matched to patients requiring tissue grafts. But to suggest that it does not involve the creation of embryos is misleading.

In 1990, the British parliament was persuaded in favour of allowing a limited range of research projects on pre-gastrulation embryos by arguments that such embryos have not progressed to the point that they can be considered as individuals. Exactly the same logic can be applied to the current ES-cell debate. ■

Can the leopard change its spots?

With less politics and more money, Italy's scientific community could take its rightful place among the world's élite.

Italy's research organizations are midway through a major process of reform, designed to sweep away the bureaucracy and deal-making that have beset them for decades (see page 264). But Italian scientists retain a degree of scepticism. They like to quote from *The Leopard*, Giuseppe Tomasi di Lampedusa's classic tale of late-nineteenth-century Sicilian politics: "Change everything so that everything remains unchanged."

If Italy's overpoliticized and underfinanced science base really is to change, two obvious things are needed: less politics and more money. In this context, 'politics' means the academic wheeling and dealing that in Italy is institutionalized by an overdose of democracy. Important decisions in Italian academia, including the distribution of research funds and senior appointments, are in the hands of committees directly elected by the academic communities they represent. The result, unfortunately, has been a system in which votes are frequently traded for favours, and which fails to reward excellence.

The reforms have successfully de-democratized a few key academic committees. For example, a new committee that distributes small grants to individual university researchers is now composed of

a handful of top academics appointed by the research ministry. It works well, concentrating funding in the best labs.

But this is still not the case for most other committees. That is why the whiff of the leopard still lingers over Italy's system of academic promotion, where most members of the new-styled selection committees are still elected, and around the corridors of the CNR, the national research council.

Italian science has an important contribution to make. Statistics published last month by the European Commission show that, although the number of scientific publications out of Italy is well below the European average when computed per million population, the number of highly cited papers exceeds the European average when expenditure is factored in.

The commitment of Italy's new centre-right government to the scientific reforms instituted by its predecessor is unclear. But if education and research minister Letizia Moratti can wean the scientific community away from inappropriate elections, and back successful reform with investment compatible with Italy's status as a G8 economic power, the rewards could be great. ■





Dome comforts
Wellcome Trust
seeks home for
biotech centre
p258



Star-gazing
Congress promises
physical sciences a
brighter future
p259



Green and mean
More research
needed as killer algae
hit Californian coast
p260



Virus returns
Chicken flu back in
Hong Kong after
ban is lifted
p261

Britain seeks transgenics deal to fend off transatlantic trade war

David Dickson, Bangkok

Britain is expected to offer this weekend's meeting of the group of eight (G8) leading economic powers a set of proposals aimed at heading off a transatlantic trade war over genetically modified (GM) food.

The measures are being fine-tuned by senior British officials after an international meeting of 300 specialists in transgenic technology in Bangkok last week. British Prime Minister Tony Blair is expected to put them to his fellow G8 leaders at their annual summit, which starts on 20 July in Genoa, Italy.

Blair may suggest greater harmonization on guidelines for measuring the environmental impact of GM organisms, and more international discussion on the matter to be held between both governments and non-governmental organizations.

Another proposal could be for an accelerated programme of 'capacity-building activities' in biosafety, particularly in developing countries.

The package is also likely to include a call



See no evil: protesters in Bangkok signal their doubts over biosafety and genetic modification.

for greater commitment to publicly funded research into the safety of the technology. Blair may argue that too much of the research on which biosafety decisions are currently based is conducted in the private sector.

All of these issues were raised at a three-day conference in Bangkok on 10–12 July,

organized jointly by the British government and the Organization for Economic Co-operation and Development (OECD), with the support of several intergovernmental organizations.

The meeting brought together participants from both developed and developing countries, representing the views of industry, government agencies, consumer and farming organizations, and environmental pressure groups.

But rather than reflecting a consensus among the participants — which the conference unsurprisingly did not find — Blair's proposals will be based on the personal conclusions of the chairman of the meeting, Lord Selborne, a former chairman of the UK Agriculture and Food Research Council.

At the conference itself, representatives of the agrobiotechnology industry, the United States, and some developing countries including China, expressed strong support for GM crops.

Jikun Huang of the Chinese Academy of Sciences said that a survey of farmers in China using *Bt* cotton, which produces a natural insecticide, had revealed a range of advantages, from increased yield to significantly lower pesticide poisoning.

But environmentalists rejected this optimistic picture and Chebet Maikut, chairman of the Uganda National Farmers' Association, said that many small farmers are worried about the extent to which GM crops

Smithsonian head ruffled by inquiry

Corie Lok, Washington

A government agency is investigating a private collection of artefacts held by the secretary of the Smithsonian Institution, Lawrence Small, to see if they were taken from endangered species.

The US Fish and Wildlife Service began an earlier investigation into Small's collection of masks, head-dresses and costumes of South American origin after pictures appeared in the January 2000 issue of *Smithsonian*, the institute's monthly magazine. Officials at the service apparently suspected that the pictures included feathers from protected bird species and teeth from endangered cats.

The original investigation ended earlier this year after Small provided import permits and a letter detailing the 1988 purchase of the collection from an undisclosed seller. A spokeswoman for the

Fish and Wildlife Service declined to give the reason behind the reopening. A spokesman for the Smithsonian said that Small is cooperating with the investigation.

Trading of items from endangered species violates the 1973 Endangered Species Act, the Convention on International Trade in Endangered Species and other federal laws. Penalties include confiscation, fines and even imprisonment.

The investigation is not the only controversy Small has faced. His proposal to streamline research at the Smithsonian by closing two research centres was sharply criticized by scientists. Earlier this year, Small reversed his decision to close one, the Conservation and Research Center in Virginia. And a Senate budget subcommittee recently voted to maintain funding for the other, the Smithsonian Center for Materials Research and Education in Maryland. ■

were allowing foreign-owned companies to control access to seeds.

Selborne said at the end of the meeting that it had not been designed to reach agreement on the issues discussed. He nonetheless drew several personal conclusions, which will be submitted to Blair as the prime minister prepares for the G8 summit.

One of these will urge governments "to ensure that there is an acceptable balance between public and private research funding for biosafety applications".

A separate report on the Bangkok meeting is due to be presented to the G8 summit by OECD officials. The OECD was asked for guidance at the G8 meeting in Cologne two years ago, and helped to organize last week's conference and a similar meeting held jointly with the British government in Edinburgh last year (see *Nature* **404**, 112; 2000) to help to formulate its response.

One proposal that emerged from Edinburgh — the creation of an international panel to study the scientific basis of concern over the safety of GM crops — was rejected by the G8 nations in the face of opposition from the United States.

British officials hope that a more limited set of proposals will win the endorsement of the G8 countries and thus help to prevent differences between Europe and the United States from developing into a full-blown trade conflict. ■

Millennium Dome sees signs of a Wellcome break

Peter Aldhous, London

London's Millennium Dome could be transformed from a national embarrassment into a leading centre for biomedical research.

The Wellcome Trust, the world's largest medical research charity, has reportedly earmarked £300 million (US\$420 million) for a bid to convert the riverside site into laboratories and an exhibition space featuring the trust's work.

The trust would not comment on the reports, which first appeared in *The Sunday Telegraph*, but issued a statement saying that its "investment portfolio is constantly under review and at any one time several opportunities may be under consideration".

The Dome was conceived as the centrepiece of Britain's millennium celebrations, its huge tent-like structure housing a series of exhibits illustrating contemporary life. It cost more than £600 million, but attracted only about half of its projected 12 million visitors during its year-long opening. The whole site now lies empty, costing the government £1 million per month to maintain.

Buying the Dome is well within the financial reach of the Wellcome Trust, which has assets of some £15 billion. The £300 million is reported to cover both the costs of purchasing the site and building laboratory facilities.

The Wellcome Trust has previously invested large sums in research facilities, most notably to build the Sanger Centre at Hinxton near Cambridge, responsible for sequencing one-third of the human genome.

But the trust has been frustrated by the denial of planning permission for a £100-million extension at Hinxton to provide space for spin-off biotechnology companies (see *Nature* **400**, 803; 1999). The trust says it still intends to pursue a revised application.

There is a shortage of laboratory space in the British capital, as highlighted by a group called the London Biotechnology Network, which said last month that up to 40 new biotech companies will need more than 5,000 square metres of lab space within the next two years. ■



Sales pitch: will Wellcome buy London's Dome?

Simple technology could reclaim America's lost votes

Mark Schroppe, Washington

Pencil, paper and simple optical scanners are the most reliable proven technologies for avoiding another voting fiasco akin to last November's US presidential election, say some of the country's leading scientists.

David Baltimore, president of the California Institute of Technology, and Charles Vest, president of the Massachusetts Institute of Technology in Cambridge, jointly launched the Caltech-MIT Voting Technology Project a few weeks after the election.

The project's findings — issued on 16 July — are that high-technology solutions have a poor record in improving voting accuracy, and that authorities should abandon them. The report, *Voting: What Is, What Could Be*, says that paper ballots and optical scanners are the most accurate, with an error rate of around 1.5%, whereas punch-card machines and electronic devices register nearly twice as many errors.

The report points out that Internet voting and other advanced solutions could

open the door to large-scale, systematic voter fraud in the United States. It argues, nonetheless, that "properly developed" high-technology options could eventually prove superior to what is currently available.

The widely publicized problems that dogged last year's vote in Florida, where George W. Bush beat Al Gore by a tiny margin, were rampant throughout the country, the project finds.

"The problem is more serious than we had imagined even last December," says Baltimore. "The voting process has simply not been taken seriously as a central element of the democratic process."

The group estimates that between 4 million and 6 million votes were lost because of ballot, registration and other problems. To slash that number by about 1.5 million in the next election, the project recommends abandoning existing electronic, punch-card and mechanical-lever systems in favour of hand counting and optical scanning of hand-marked ballots. It also calls for a new system for field-testing

other alternatives.

According to the report, procedural changes such as making regional registration information available at polling sites could prevent the loss of another 2 million or so votes, but further reductions will require better voting technology.

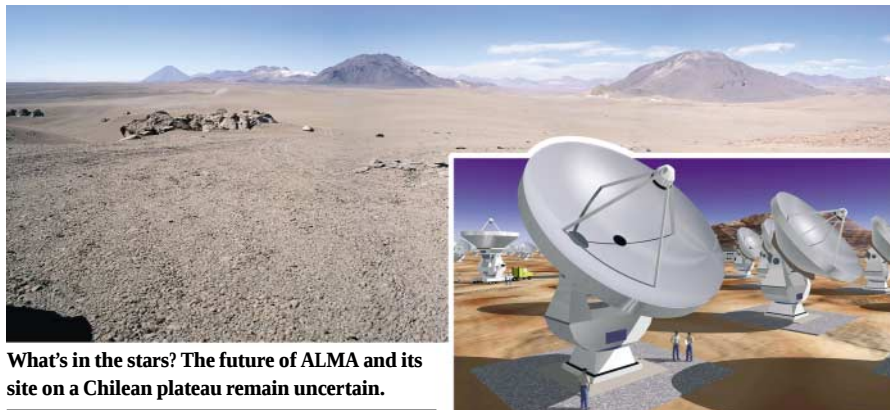
It recommends federal funding to develop new voting technologies which could include ballots, either electronic or paper, that can be printed on-site, allow voters to confirm their selections, and be stored for later review. Although optical-scanner systems come closest to that, they have drawbacks, including high costs, inflexibility and significant scanning inaccuracy.

"We want every American's vote to count," says Vest. To that end, he says, the next phase of the project will focus on the design of new equipment and processes. "Computer technology appropriately applied offers significant opportunity for cost-effective advances," he says. ■

♦ <http://www.vote.caltech.edu>

Congress touts budget boost for NASA and the NSF

ESO



What's in the stars? The future of ALMA and its site on a Chilean plateau remain uncertain.

Tony Reichhardt, Washington

One small step in the federal budget process last week was a giant leap forward for the funding prospects for non-biomedical research next year in the United States.

A key congressional subcommittee voted on 10 July to boost funding for the National Science Foundation (NSF) and NASA for the 2002 fiscal year, which begins on 1 October. The move is the latest and largest of a series of congressional actions to restore research funds that were omitted from President Bush's budget request in April (see *Nature* **410**, 731; 2001).

Scientists have been grumbling for months about the White House's proposed 1.3% increase for the NSF. Now the House subcommittee that oversees the foundation's budget says that the increase should be 9.5% instead, bringing the total NSF budget to \$4.84 billion.

The full House appropriations committee was due to take up the measure this week, as was a Senate appropriations panel chaired by Barbara Mikulski (Democrat, Maryland). Mikulski is a supporter of NASA and has said she favours increased funding for the NSF.

Budget deliberations will continue into September, when the House and Senate will reconcile their spending bills. But the marker laid down by the House panel augurs well for both agencies' prospects, their supporters say.

Every directorate in the NSF saw its allocation raised by the subcommittee. And at least one project — the High-performance Instrumented Airborne Platform for Environmental Research (HIAPER), an aircraft for conducting atmospheric science to be operated by the National Center for Atmospheric Research in Colorado — was saved from elimination. The House panel added \$35 million for the project, which had been given nothing in Bush's request. Funding for Major Research Instrumentation, which supplies university researchers with

advanced lab equipment, was restored to last year's level of \$75 million, where the White House had asked for only \$50 million.

Still uncertain is the fate of the Atacama Large Millimeter Array (ALMA), a 64-antenna telescope planned to be operational by 2010. The United States, Europe and Japan expect to spend a combined \$600 million on ALMA, which will be located in Chile. But the NSF, which is funding the project on the US side, has been slow to commit construction funding for it. Astronomers were alarmed this spring when the White House moved funds for ALMA out of the NSF's Major Research Equipment account — traditionally where large construction projects are handled. The House panel moved the funds back, but kept ALMA's budget at \$9 million. This means it would remain stuck in the design phase — unless the Senate adds more funds.

NASA got a raise from the subcommittee that was \$415 million over the presidential request, bringing the total for 2002 to \$14.9 billion. But the space agency's biggest fiscal headache, the International Space Station, remains unresolved. The House panel added \$310 million to the White House request for the station, bringing it to \$1.8 billion. Most of that would go to resurrect work on a crew return vehicle, which Bush scrapped but which many scientists say is needed to support an active research programme.

But the rise hardly makes a dent in the cost overrun — \$4 billion and growing — that prompted NASA managers to propose deep cuts in research aboard the station (see *Nature* **410**, 399; 2001). Life scientists planning biological studies are aghast at NASA's plans to cancel key pieces of lab equipment in order to trim its research budget. Among the casualties are facilities for holding animals, plants and cell cultures. In a draft report accompanying its recommended budget last week, the House panel seemed as frustrated as the scientists, and has begun an investigation into the cost overruns.

Physical sciences lose ground as US shifts towards biology

Jonathan Knight

The number of graduate students enrolling in physical sciences in the United States is still falling sharply, despite an expansion in the government's overall research budget, according to the National Research Council (NRC).

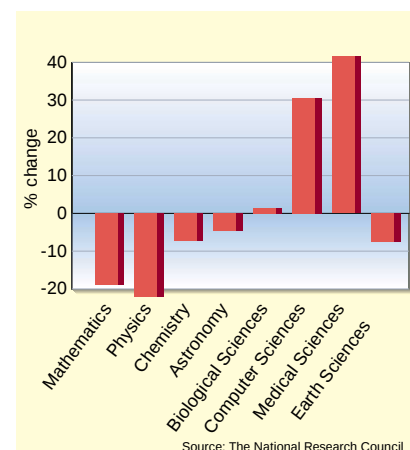
In its latest study, the NRC finds that recruitment has dropped in chemistry, physics, mathematics and engineering as federal funding shifts towards the life sciences. For example, the number of full-time graduate students in physics fell by 22% between 1993 and 1999, whereas the number in medical sciences rose by 41%.

Some observers expected increases in overall research funding to stop or slow the decline, says Stephen Merrill, staff director of the NRC study. "There was an expectation that the rising tide was lifting all boats," he says.

But an NRC panel chaired by Dale Jorgenson, an economist at Harvard University, found that the rise in total research investment after 1997 had failed to reverse the overall trend. The panel noted that federal funding for physical sciences declined by 18% in constant dollars from 1993 to 1999, whereas funding for life sciences rose by 28%.

Michael Lubell, a physicist at the City College of New York and head of public affairs at the American Physical Society, predicts that effects of the shift will be felt gradually, as the quality of the workforce declines. "As researchers retire from the national labs and academia, it will get harder to replace them with high-quality people," he says.

♦ www.nationalacademies.org/step



Percentage change in the number of full-time graduate students, United States, 1993–99.

Search for alien life reasserts its credibility

William Triplett, Washington

The much-maligned Search for Extraterrestrial Intelligence (SETI) programme is slowly rebuilding its credibility in Washington, where some law-makers would like to see NASA restore links with it.

Congress stopped all public funding for SETI in 1994, after some members ridiculed it as squandering taxpayer's money on a quest for "little green men". But the project has made considerable headway since then with private funds, and on 12 July a congressional committee heard testimony on it, as part of an overview hearing on research into whether life exists beyond Earth.

Leaders of the project, which is organized by the SETI Institute in Mountain View, California, say they are not seeking a restoration of government funds. But Christopher Chyba, one of the institute's directors, told

the hearing of the space and aeronautics subcommittee of the House Science Committee that the project would like to overturn the current ineligibility of SETI proposals to compete for peer-reviewed grants from NASA's astrobiology programme.

"These grants should be open to SETI researchers to apply for in a peer-reviewed way just as they are for anyone else," says Chyba. He adds that SETI, by searching for other intelligence, is addressing one of astrobiology's central questions.

Since losing its federal funding in 1994, SETI has not only survived but has grown through philanthropic donations, primarily from wealthy technology pioneers such as William Hewlett, David Packard, Gordon Moore, Paul Allen and Barney Oliver.

Michael Meyer, a senior scientist at NASA's astrobiology programme, says that



as a result of the congressional language used in removing SETI from the federal budget, the agency is effectively prohibited from funding SETI-specific activities even through grants, which are awarded only to proposals involving "good science".

SETI critics have argued in the past that SETI does not constitute good science. But Lamar Smith (Republican, Texas), whose enthusiasm for the project led to the hearing taking place, noted that SETI's track record has improved markedly in recent years.

Indeed, Chyba points out that successive decadal reviews of astronomy by the National Academy of Sciences have endorsed SETI. The most recent one singled out the project for pioneering new technology that will have other useful applications.

But whether NASA is ready to offer astrobiology grants to SETI is another matter. "The SETI Institute has been very successful and they do some good stuff," says Meyer. "But there's the issue of once burnt, twice shy. We're growing an astrobiology programme that seems to be fantastically popular with not only the public but with the White House Office of Management and Budget and with Congress and even within NASA. But there could be concern that if we include SETI in this, while intellectually it fits within the programme, all of a sudden the enthusiasm you had gets turned back into why are we looking for little green men?"

But Smith told the hearing: "The discovery of life in the Universe would be one of the most astounding discoveries in human history. Funding should match public interest and I don't believe it does." He was backed up by Zoe Lofgren (Democrat, California). According to one congressional staff member, Congress is moving slowly towards an acceptance, if not an embrace, of SETI. "There's really no groundswell for restoring a federal funding line for SETI right now," says the staffer. "Then again, there's no real opposition to it, either."

♦ <http://www.seti-inst.edu>

Action urged to combat killer algae

Rex Dalton, San Diego

More than a year after invasive algae took root along the southern California coast, authorities have yet to arrange a thorough scientific study of either the algae or the programme to eradicate them.

At a workshop held on 10–11 July in San Diego, a sub-panel of the US Aquatic Nuisance Species Task Force agreed that such studies were urgently needed to combat *Caulerpa taxifolia*, which has infested two lagoons in southern California.

"Research and peer review must go hand-in-hand with eradication and control efforts," says Steven Miller, director of the National Undersea Research Center at the University of North Carolina and chair of the panel, which is preparing a national action plan to fight future infestations.

C. taxifolia is indigenous to the Caribbean, but a strain of the algae has already destroyed large areas of natural habitat in the Mediterranean after it was inadvertently released near Monaco in 1988. The fast-growing algae are toxic to some marine life and have no natural enemies in the areas they have invaded.

In California, the algal growths — some estimated to be as large as 30 tonnes — were identified in June last year (see *Nature* 406, 447; 2000).

The source of the algae is unknown, but the authorities suspect that they were dumped by aquarium enthusiasts who use them in their tanks. A bill is moving through the California legislature to outlaw the sale and possession of *C. taxifolia* and eight other similar algae, but it has faced opposition

from the home aquarium business.

Susan Williams, director of the Bodega Marine Laboratory at the University of California, Davis, has spent the past year pressing for research and outside peer-review of the eradication effort. She says she has been unsuccessful because of resistance by some state and federal officials who want to concentrate resources on eradication.

The \$1-million eradication effort involves placing tarpaulins over patches of algae and injecting chlorine underneath to kill them. There is evidence that the chlorine has killed algae under the tarpaulins, the workshop was told. But patches of the algae have also cropped up outside the tarpaulins. ■



Green death: *Caulerpa taxifolia* threatens to smother parts of the Californian coast.

Outbreak of chicken flu rattles Hong Kong

David Cyranoski, Hong Kong

Despite a huge poultry cull, Hong Kong is struggling against an outbreak of a potentially deadly chicken flu virus.

The virus, H5, normally lives in the intestines of aquatic birds. Although it cannot infect mammalian intestines, it can attack the respiratory tract, causing flu-like symptoms and, in extreme cases, pneumonia. In 1997, the H5N1 strain killed six people in Hong Kong.

Although the H5 strains found in Hong Kong this year are not thought to be transmissible to humans, the authorities acted swiftly to contain the outbreak. In May, around 1.3 million fowl were killed, including almost the entire stock of chickens, and imports were banned.

But on 5 July, only three weeks after the ban was lifted, traces of a related virus were found in fecal samples from a chicken imported from China. Researchers are now sequencing the genome of the virus, in an effort to determine its identity and its disease-causing potential in both chickens and humans.

The May slaughter, which cost the Hong Kong government HK\$245 million (US\$32 million) in compensation, is seen as overkill by some researchers. "The reaction to the virus was completely overdone," says a food hygienist at one Hong Kong university, who did not want to be identified.

But Nikolaus Sucher, a molecular biologist at Hong Kong University of Science and Technology, believes the decisive action may have prevented a serious international epidemic.



Culling spree: Hong Kong health officials have killed 1.3 million fowl in an attempt to beat the flu.

Ken Shortridge, a microbiologist at Hong Kong University, agrees that the H5 strains detected this spring were not transmissible to humans. But he says that the viruses were still a threat because they can swap genes with related viruses in other fowl, which could result in a virus that can infect humans. "All of the ingredients for generating a virus like the deadly one in 1997 were in the poultry markets this spring," Shortridge says.

In the wake of the outbreak, Hong Kong has stepped up sanitation rules, stipulating that markets be disinfected once a month and that chickens be segregated from other fowl.

The government is now considering using different tests to detect H5. In the past, it

relied on testing for a bird's immune response to the virus. But now DNA-based techniques have been developed by a local biotechnology company, Hong Kong DNA Chips.

Government scientists who control the testing claim that the DNA tests cannot replace immune tests, and have confined their use to confirming results. But the DNA-based tests are quicker, and the government is under pressure to use them more widely.

Most of Hong Kong's chickens come from mainland China, which raises the politically sensitive question — on which no-one *Nature* interviewed would comment — of whether inadequacies in China's export regulation are allowing the virus into the territory. ■

Disgruntled homeopaths seek remedy in court

Alison Abbott

A leading Italian television presenter is being sued by two homeopathy organizations for failing to include their views in a broadcast of his prime-time science programme *Superquark*.

Millions of Italians use homeopathic products — which purport to treat diseases using vanishingly small doses — and the homeopaths claim that the broadcast was unfair and could threaten their business. They complain that Piero Angela, who produces his own show, selected only interviewees who were critical of homeopathy for a show that was broadcast last July.

But Angela has received strong backing from the scientific community. He says that he has received many letters of support from individual researchers, including Nobel prizewinners Renato Dulbecco and Rita Levi-Montalcini.



Piero Angela's broadcast has been backed by scientists.

Angela says that Italian public television archives show that that 14 times more programmes had been broadcast advocating homeopathy than criticizing it. His programme sought to redress the balance, he says. He also argues that it is not his job to tell viewers what they

want to hear: "Science is not like philosophy, where viewers can listen to both sides and decide for themselves," he says. "Science cannot be decided on by the vote of viewers."

One of the guests on Angela's show was Antonio Cassone, head of the department of bacteriology at the Istituto Superiore di

Sanità, Italy's national health institute, who is a member of the ministry of health's ad hoc committee on homeopathy.

On the show, Cassone says, he merely offered the opinion that safety information should be provided on all homeopathic products, and that efficacy information should be provided on anything that is to be injected. "I would have had no objection to a homeopathist sharing the show," he says, "because it would have been even more convincing to the viewers that arguments of homeopaths against providing information are untenable." But he fully supports the way Angela produced his programme.

The suits, one civil and one criminal, have been brought by the Catania-based Italian Association of Medical Homeopathy and the Rome-based Italian Federation of Associations of Medical Homeopathy. The cases should come to court in the autumn. ■

Frustrated stem-cell pioneer quits United States for Britain...

Washington A leading US cell biologist is to relocate to Cambridge University, citing frustration with political obstacles to his work in the United States.

Roger Pedersen will leave the University of California, San Francisco, to take up a faculty position at Cambridge, a leading centre in the field. Human-embryo research, including work on stem cells, is legal and publicly supported in Britain.

Federal funding for research in which embryos are destroyed has been banned in the United States since 1996. Pedersen's work on stem cells has been funded by Geron, a Californian biotechnology company.

Pedersen told *The Wall Street Journal* that he faced a choice between "very favourable circumstances and tremendous support" in Britain and "sitting on my hands for the next few years" in the United States.

...as UK universities fall foul of retiring Harvard president

London The past president of Harvard University has cast aspersions on the state of British higher education, telling a group of

students that it is turning from "a disaster into a nightmare".

Neil Rudenstine, who has just retired from his Harvard post, said: "The sad reality is that in Britain even Oxbridge is suffering because of the lack of resources and private-sector involvement." He added that Cambridge's famous Cavendish laboratory had a budget of US\$16 million which, he said, "wouldn't keep our history department going".

The off-the-cuff remarks by Rudenstine, who was a Rhodes Scholar at Oxford and obtained a degree from New College in 1959, were reported by a British journalist on a scholarship at Harvard. A spokesman for Harvard University said that the remarks were made privately and would have been phrased differently if Rudenstine had known they were going to be reported.

Cycling officials aim to avoid Tour de Farce

Paris Organizers of this month's Tour de France cycling race have called on scientists to help them to win the battle against doping.

A spate of highly publicized doping scandals has hit recent Tours, and many believe that drug-taking is rife throughout professional cycling. As part of a programme of antidoping measures aimed at cleaning up the three-week-long Tour, the organizers



Uphill struggle: organizers of the Tour de France are working hard to win the war against doping.

have signed a three-year research contract, worth FF1.5 million (US\$200,000), with the CNRS, France's main basic research agency.

The research will focus on the physiological effects of intense physical activity, individual and social vulnerability to doping and the physiology of high-level sporting achievement.

Voluntary ban agreed on reproductive cloning

Tokyo The International Congress of Developmental Biology has called for a worldwide moratorium on cloning for reproductive purposes, as scientists aim to dissociate themselves from those who

want to clone humans.

The congress, which met in Kyoto, Japan, last week, was told that mammalian embryos produced by cloning often have fatal defects. Some researchers had hoped that the moratorium call would be accompanied by a statement supporting cloning for therapeutic purposes, but no consensus was reached.

Several countries, including Japan, have declared legal moratoria on cloning, but this is thought to be the first such announcement by an international group. Walter Gehring of the University of Basel, president of the International Society of Developmental Biologists, said that the statement would help to ensure self-regulation by scientists.

Physicists cling to lure of the public sector

Paris Staff at a synchrotron near Paris have been on strike for several weeks, in protest at the government's refusal to secure their public-sector status after the facility is replaced in 2005.

The existing Lure synchrotron is to be replaced by a FFr2.4 billion (US\$312 million) facility, known as Soleil, that will be paid for by the government but operated by a private company. The government says the new arrangement will be more flexible, but

unions representing the staff say that public-sector status would allow staff more input into management decisions and greater job security.

Union representatives will decide this week whether to continue the strike, after a meeting with the management of the CNRS, France's main basic research agency, which operates Lure and will be the majority shareholder in Soleil.

Tide of argument breaks at ocean meeting point

Cape Town Local scientists have been dragged into a feud between two South African towns over the joining point of the Atlantic and Indian Oceans. The truth, say oceanographers, is that no such meeting place exists.



Water fuss: tourism is at the centre of a row over the meeting of the Atlantic and Indian Oceans.

Cape Point, which is near Cape Town, generates more than US\$4 million a year from visitors coming to see the meeting point between the two oceans. The town's claim has some support from biologists, as it is the place where flora and fauna from the east and west coasts of Africa meet. But oceanography textbooks favour Cape Agulhas, which is on the southernmost point of Africa, because currents from the east and west mix there. Cape Agulhas, which does not attract tourists, has complained to the public prosecutor about Cape Point's claims.

But oceanographers say that the line is impossible to define, as the currents constantly shift. "If you think of two buckets of water being poured together, there's no dividing line between them," says Anthony Richardson, an oceanographer at the University of Cape Town.

Correction The European Space Agency's Planck satellite, scheduled for launch in 2007, will measure the polarization of the cosmic microwave background radiation to an angular resolution of 10 arcminutes or better, not 10 degrees as was stated in the News Feature of 21 June (see *Nature* **411**, 880–881; 2001). *Nature* apologizes for this error, which was introduced during editing.

Forza scienza!

Italy is a major economic power, but it underachieves in research. Alison Abbott examines attempts to reform the nation's scientific institutions — and considers their prospects under the new government of Silvio Berlusconi.

As Italy hosts the G8 economic summit in Genoa this week, its scientists can be excused for reflecting on the irony of their position. They work in one of the world's wealthiest countries, yet on most measures of activity in scientific research, Italy compares not with powerhouses such as the United States, Germany or Japan, but with such minor scientific nations as Portugal (see figure, opposite).

The problems of Italian science are partly a result of chronic underinvestment. But decades of bureaucracy and cronyism must share the blame. Until recently, with academic appointments placed in the hands of powerful centralized committees, patronage held sway. The limited sums made available for research, meanwhile, were spread thinly across disciplines and research groups, with little attempt to promote priority projects. The problems of Italian science, in fact, were shared by the entire public sector, allowed to fester by a corrupt political élite.

But in 1997, after the 'clean hands' judicial investigation into corruption had swept away the old political guard, the centre-left



In safe hands? Research and education minister Letizia Moratti with Prime Minister Silvio Berlusconi.

'Olive Tree' government led by Romano Prodi — now president of the European Commission — began to reform the system. In science, those reforms are still a work in progress. And with the Olive Tree coalition cut down in April's general election by the challenge of the populist media tycoon Silvio Berlusconi, their future is unclear.

Unknown quantity

Basic science seems unlikely to be a priority for Berlusconi's centre-right coalition: the campaign of his *Forza Italia* (Come on Italy!) party had virtually nothing to say on the subject. The head of the new 'superministry' for education and research, Letizia Moratti, is an unknown quantity in scientific circles — she previously headed an investment house and served briefly as president of RAI, Italy's state television network. Her deputy, nuclear engineer Guido Possa, was one of Berlusconi's schoolmates. "I'm waiting and watching to see what happens," says developmental biologist Edoardo Boncinelli of DIBIT, the Department of Biological and Technological Research at Milan's San Raffaele Institute.

The Prodi government kick-started its reforms with a law pushed through in 1997 by public works minister Franco Bassanini. For the following two years, this allowed existing statutes governing state institutions to be revised without parliamentary approval. Research organizations were subjected to the Bassanini reforms. But the details, and many key appointments, are still being filled in.

Attempts to restructure the CNR, Italy's national research council, which supports research in fields from psychology to chemical engineering, illustrate the current situation. A decree passed in April 1999 proposed a thorough restructuring, condensing the

CNR's 330 institutes and centres to 100, mostly through mergers designed to achieve a critical mass in each research area.

It also sought to introduce a simpler organizational structure. As a first step, two years ago, the CNR's 15 discipline-oriented 'advisory' committees were dissolved. These committees, dominated by university professors, exerted tremendous power. Not only did they control research funds earmarked for the CNR, but they also decided on appointments in its institutes and influenced the direction of large strategic projects funded by the research ministry. They distributed funds thinly and widely in an attempt to keep everyone reasonably happy, rather than rewarding excellence and innovation.

Power over the CNR's spending and administrative decisions will now rest in the hands of a nine-member executive council known as the *consiglio direttivo*. Four of its members will be appointed by the Assembly of Science and Technology (AST), which will also advise the government on all aspects of science policy. A further four will be elected by academic researchers, with the ninth being the CNR's president, Lucio Bianco. But with the AST still awaiting appointment, and with methods for electing academic representatives onto the CNR central committee still to be agreed, the council so far exists only in interim form.

The huge task of restructuring the CNR's institutes is being carried out, albeit slowly, by this interim council. The directorships of half of the new institutes have been advertised internationally — previously they were not even advertised in Italy — and the first handful of selections is expected to be made within the next few weeks.

The CNR's 4,000 or so researchers seem

STEPHANIE MAZECORRIS



Life's speeding up at universities such as Bologna.

reasonably happy with the changes made so far. "What is clear already is that the CNR reforms are very real," says Alberto Passerone of the CNR Institute for Physical Chemistry of Materials in Genoa. "The impacts will be beneficial, in my opinion."

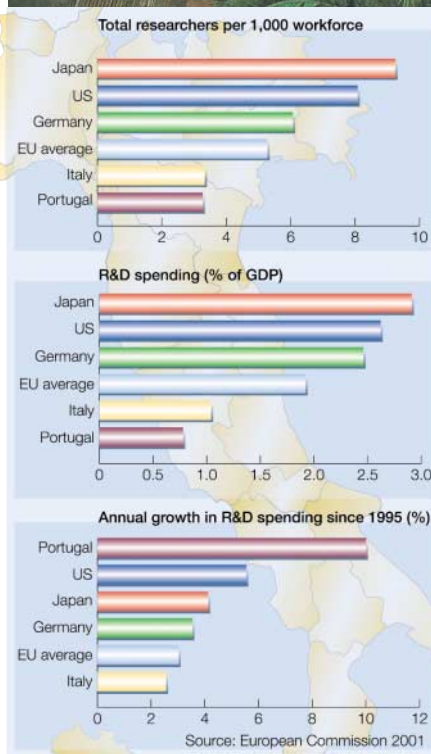
At Italy's smaller research agencies, reforms are progressing with varying degrees of success. The INFN, which supports research into nuclear and high-energy physics, was always efficiently run, and so has escaped major changes. Scientists funded by the Italian Space Agency (ASI), meanwhile, have profited from the reform process. ASI has no research institutes of its own, but its new governing statute, passed in 1999, made research its first objective. Since then, ASI's annual research budget has more than trebled, to 200 billion lire (US\$87.5 million), not counting Italy's contributions to the European Space Agency's science programme. "It is more than we have for our national programme in Germany," says an envious Günther Hasinger, director of the Max Planck Institute for Extraterrestrial Physics in Garching.

But the 1,500 staff of the ISS, the Higher Institute of Health in Rome, fear that the reforms have made a bad situation worse. The ISS both conducts medical research and is responsible for technical aspects of drug regulation. It was supposed to be reformed to boost its efficiency and make it independent of the health ministry. But its new statute, finalized earlier this year, appears to give even more power to bureaucrats who have traditionally frustrated the working lives of ISS scientists. "The situation is worse than ever," says one disgruntled staff member. Worryingly, research is not even mentioned in the ISS's new mandate. "We won a few battles in the restructuring, but we lost the war," says former ISS director Giuseppe Benagiano. Since April, the revamped ISS has been headed by Enrico Garaci, a microbiologist and former president of the CNR, who is seen as an old-style political appointee.

Wheeling and dealing

The research agencies are only half of the picture, however. The universities have contributed to the problems of Italian science, in particular through a system of academic appointments based on national competitions, or *concorsi*, that were notorious for the wheeling and dealing conducted on the appointment committees. During the mid-1990s, state prosecutor Adelchi D'Ippolito brought 50 cases alleging corrupt practices during *concorsi* — although none has yet led to a conviction.

In 1998, the national *concorsi* were replaced with a system that allows individual universities to make their own appointments. Dario Braga, a chemist at the University of Bologna who argued for reform for more than a decade, says that the advantage



of the new system is speed. "The old *concorsi* procedures used to take ages to draw conclusions, and the value of speed in Italy should not be underestimated," he says.

But he is disappointed that deal-making is still a large part of the process. Appointments at each university are now made by five-member committees, four of whom are elected at the national level by the relevant academic community. These elections are frequently rife with deal-making, the goal being the selection of committee members who will favour the 'right' candidate. These are often internal candidates who are cheaper to hire. The previous government recognized this problem, and provided funding to cover the additional costs of recruiting external candidates. But whether Berlusconi's administration will continue this policy is unknown.

Indeed, the new coalition's policies on science and the universities remain mysterious. *Nature* requested an interview with Moratti,

Slimmed down: bureaucracy at the national research council's base is being streamlined.

but she is not yet speaking about her plans.

Near the top of Moratti's in-tray must be the implementation of Italy's first national research plan, drawn up by the former government with the help of a high-level expert committee. It includes a series of strategic programmes, in areas such as post-genomics and nanotechnology, which are supposed to be funded over the next three years by a one-off windfall of some 800 billion lire from the sale of licences to companies operating mobile-phone networks. Researchers are now wondering whether the first call for proposals will go out in September, as planned, whether the programmes will be redirected into applied projects to reflect Berlusconi's business-oriented policies, or even whether a new set of priorities will be drawn up.

Scientists in the south of Italy, meanwhile, are nervous about the election of a government dominated by politicians from prosperous northern cities. The previous government supported investment in research in the economically depressed south, including a 40-billion-lire-development fund known as Biogem to build up a concentration of genetics in the Naples area. Biogem's president, Roberto Di Lauro, a geneticist at the Federico II University of Naples, says he is "waiting for the new government to settle, to see what commitment it will make to the south".

When it comes to the ongoing reforms that are the biggest issue facing Italian science, leading researchers want Moratti to push through the changes initiated by the previous government, and reward successful reform by increasing Italy's spending on research to levels commensurate with its status as a G8 nation. "It will be hard to operate the new system efficiently if there is no new money made available," says Arturo Falaschi, director of the International Centre for Genetic Engineering and Biotechnology in Trieste, and a member of the CNR's interim executive council. ■

Alison Abbott is *Nature's* senior European correspondent.

Windows on the brain

By monitoring the activity of up to a hundred brain cells at once, neuroscientists are gaining new insights into brain function and paving the way for fully functioning prosthetic limbs. Marina Chicurel reports.

A robotic arm flexes as it receives commands from a monkey's neurons. Scientists probe an animal's perceptions and intentions. The dreaming brain comes into view as a rat's memory-traces are shown being replayed during sleep.

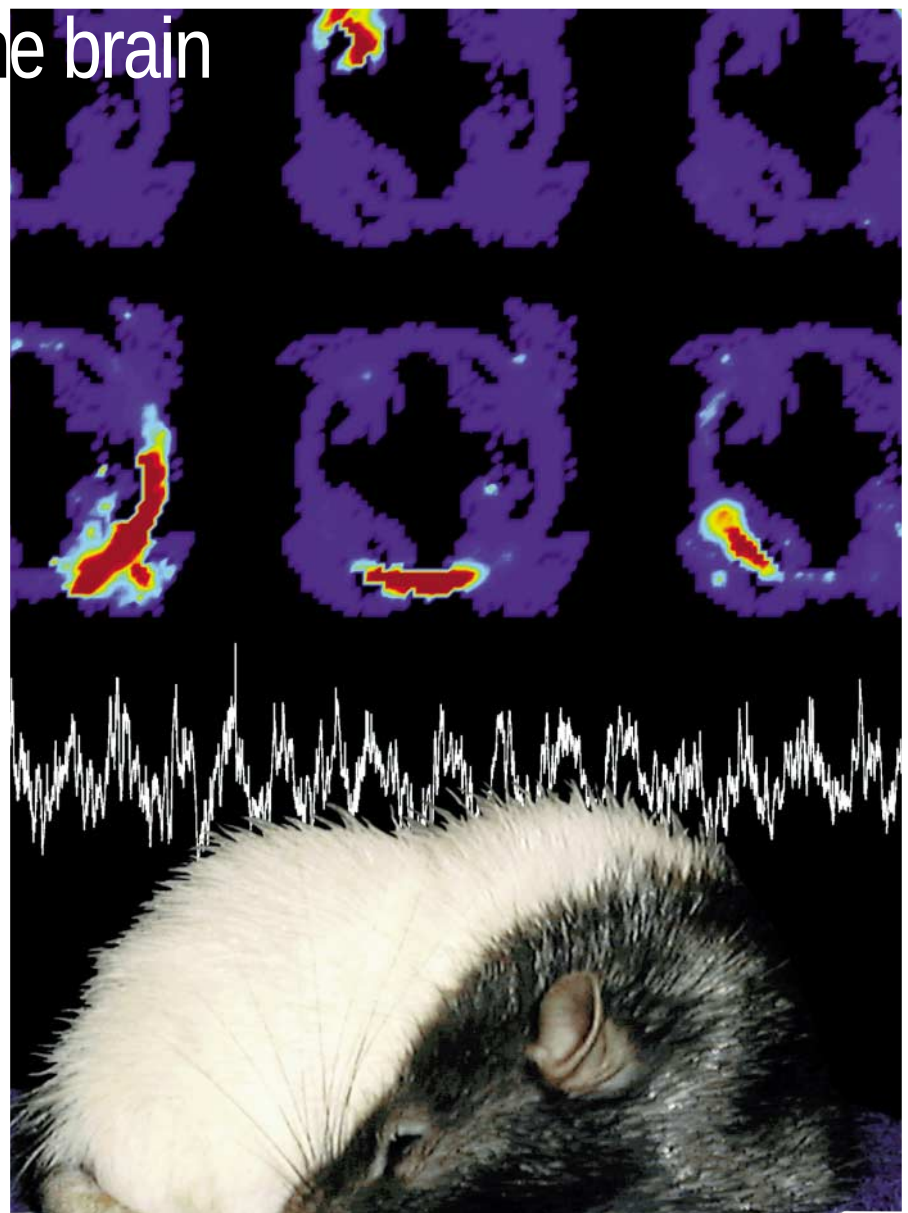
Thanks to tools developed to monitor the activity of up to 100 individual neurons simultaneously, these futuristic scenarios are now an everyday reality in some of the world's top neuroscience labs. Such 'multi-unit' recording is poised to transform brain research. "It will mark the transition of neurophysiology from small science to big science," predicts Vernon Mountcastle of Johns Hopkins University in Baltimore, Maryland, whose recordings from single nerve cells helped to underpin present textbook principles of brain function.

For decades, neuroscientists have implanted electrodes into the brain to listen in as neurons talk to one other. When one brain cell sends a signal to another, it fires off a voltage pulse along a fibre leading towards the next neuron. This pulse can be detected by electrodes placed close to the firing neuron.

Head start

Traditionally, researchers have monitored the activity of brain cells one at a time. Such studies have made huge contributions to our understanding of the brain but, ultimately, they cannot lay bare the behaviour of large, interconnected networks of neurons. Searching for a glimpse of this bigger picture, many neuroscientists are now inserting multiple microelectrodes into animals' brains.

The technique first attracted attention in



Hard day's night: patterns of neuronal firing recorded as a rat sought food in a track (top) are replayed when the rodent 'dreams'.

the early 1970s, when researchers mapping connections within the brain began comparing readings taken from pairs of neurons¹. They noticed that, in some cases, the firing of one neuron preceded the silencing of another, and reasoned that the first neuron was sending signals that, either directly or indirectly, inhibited the second cell. The technique has since been used to map the connections between many different brain areas.

But more ambitious multi-unit studies were slow to take off. "It is more difficult to record from two or more neurons than it is from one," explains Charles Gray of Montana State University in Bozeman. "If you can get a lot of information from one neuron,

why make your life more difficult?" Others in the field suggest that researchers were unsure of how to analyse multi-unit data and that many were not even certain what questions the technique could address.

In the late 1980s, Gray, together with Wolf Singer of the Max Planck Institute for Brain Research in Frankfurt, helped to frame one of those questions. Researchers already knew that the rate at which a neuron fires off pulses encodes important information. But Gray and Singer were intrigued by suggestions that the precise timing of the individual pulses might also be important. Theoretical studies² had indicated that neurons working on the same problem might fire synchronously.

Gray, Singer and their colleagues examined this idea in 1989, using up to seven electrodes to record the activities of widely

distributed neurons in visual areas of cats' brains. When the cats looked at moving bars of light oriented in a particular direction, the researchers saw that neurons specialized to respond to this orientation fired in an oscillating pattern. These oscillations were synchronized even when the cells were far apart from one another and involved in processing information from different areas of the field of view. The researchers suggested that the timing of the neurons' pulses might help to coordinate such distant cells when they are processing similar types of information³.

All together now

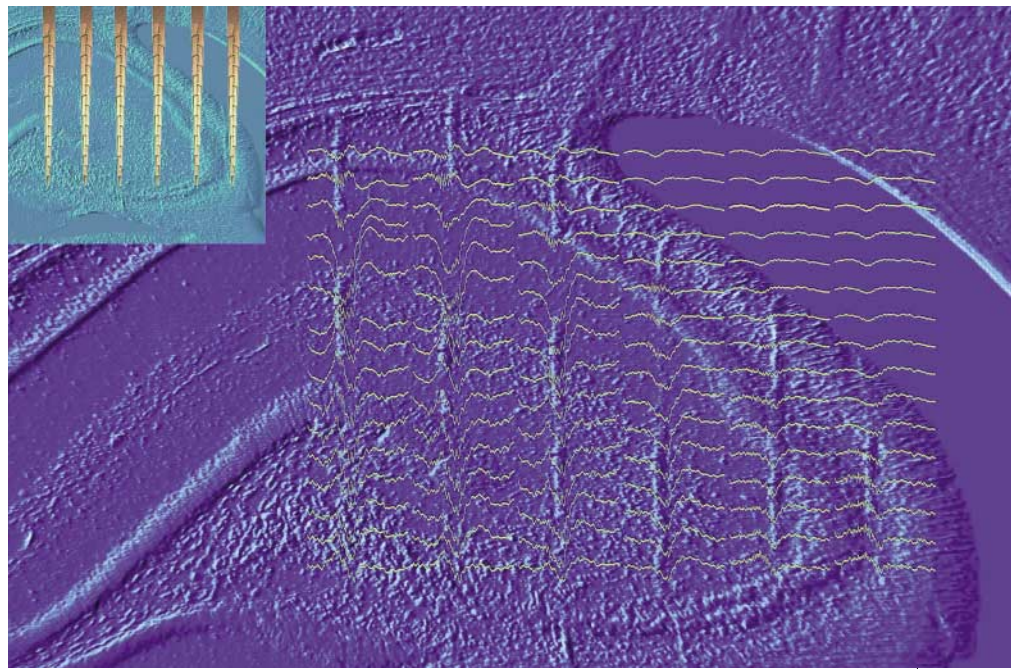
Gray and Singer's experiments motivated many others to turn to multi-unit recordings. The precise significance of the oscillations they saw remains a matter for debate. But dozens of multi-unit studies have since shown that synchronous firing is associated with visual perception and the conscious processing of other types of information⁴.

Last year, Singer's team showed anaesthetized cats a checked pattern made up of two different sets of stripes moving at right angles to each other. Varying the brightness of the stripes changes the way the overall pattern is perceived — it is either seen as two individual moving sets of stripes, or a single, checked pattern moving as a whole. Using multi-unit recordings of cells in a visual processing area of the cats' brains, Singer showed that neurons that responded to different sets of stripes fired synchronously when a cat was shown stripes that should have been perceived as a single pattern, but not when the stripes would have appeared to move as two individual sets⁵.

Building from the early experiments that used information from pairs of neurons to map connections between different brain areas, some neuroscientists are now finding that cells that are wired together may show strikingly similar behaviour — underlining the precision with which the brain's connectivity can develop. For example, researchers led by José-Manuel Alonso at the University of Connecticut in Storrs and Clay Reid at Harvard Medical School in Boston have recorded simultaneously from connected neurons in two parts of the cat's visual processing pathway — the lateral geniculate nucleus (LGN) and the primary visual cortex.

The LGN is the first relay station in the brain that receives signals from the retina, which it then passes on to the primary visual cortex. Last month, Alonso and Reid reported that the probability of finding a connection between an LGN cell and a cortical cell is highest when the pair share similarities in at least five behaviours⁶. These include the precise timing of their responses and the size of the area in the field of view that they respond to.

Multi-unit experiments can also provide insights into the workings of the sleeping brain. In a study published in January⁷, Matthew Wilson and Kenway Louie of the



Finely tuned: a 200-millisecond neuronal event captured from 96 brain sites simultaneously using ultrafine silicon probes (inset), each offering 16 independent recording electrodes.

Massachusetts Institute of Technology described how they simultaneously tracked the activities of up to 13 neurons in rats' hippocampi — a brain region involved in spatial learning and memory — while the animals ran around a track to collect food. They then monitored the same neurons as the rats slept.

Wilson and Louie spotted patterns of neuronal activity during rapid eye-movement (REM) sleep, some of which lasted for over a minute, that closely matched those recorded when the animals were collecting the food. In humans, at least, REM sleep is a phase in which dreams occur. Wilson believes the patterns he and Louie observed are 'replays' of memories formed earlier in the day, and that they may underpin the rats' dreams.

These observations may shed light on how waking experiences are transformed into memories. Wilson and many others in the field believe that animals use these

replays to process information gathered while they are awake, and to lay down long-lasting memories of the day's experiences.

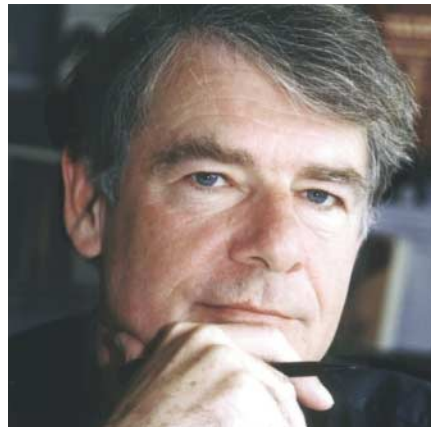
Other researchers, meanwhile, want to put the multi-unit technique to clinical use. They are using multi-unit recordings to develop devices that turn thoughts into actions — a concept that could lead to the development of functional prostheses for paralysed patients.

Out on a limb

As early as 1970, researchers showed that by monitoring the activities of just five neurons in an area of a monkey's brain controlling limb movement, they could predict how the force exerted by the monkey's wrist would change⁸. Subsequent multi-unit experiments have revealed how neurons in such 'motor' areas of the brain encode information about the direction of movement⁹, so that it is now possible to predict from such recordings how an animal will move its arm¹⁰. In the long term, researchers hope that people who have lost the ability to move their limbs may be able to control prosthetic limbs via electrodes implanted into motor areas of their brain.

Researchers have recently made progress towards this goal by using the output of multi-unit recordings to control robotic devices in real time. Last year, a team led by Miguel Nicolelis of Duke University in Durham, North Carolina, monitored neurons in regions of a monkey's brain involved in planning movements as the monkey prepared to move a joystick to the left or the right, or to reach for a piece of food¹¹.

As the monkey repeated its task, Nicolelis used a computer to develop algorithms that



Live link-up: Wolf Singer suggests that neurons synchronize themselves to perform certain tasks.



Mechanical mimic: Miguel Nicolelis' robotic arm can copy a monkey's movements in real time.

described the movement in terms of neuronal activity. In both tasks, no more than 10 minutes of repetitions were needed before the algorithm could predict, in real time, how the monkey would move its arm. For the joystick task, Nicolelis then used the algorithm to turn the signals from the monkey's neurons into commands for a robot arm, which simultaneously executed very similar movements to the monkey's own hand.

Andrew Schwartz of Arizona State University in Tempe and the Neurosciences Institute in San Diego, California, is working on the same problem using subtly different methods. "My goal is to make a movement that looks natural," says Schwartz.

He is testing algorithms that use neuronal signals to predict a monkey's arm movements in three dimensions as it interacts with objects in a virtual-reality environment. "The animal

The picture of a brain network should become crisper if more neurons can be measured.

sees a series of balls that are floating in space," explains Schwartz. "One of the balls is a cursor that's hooked up to its wrist. What the monkey has to do then is move the cursor to touch one of the other balls."

Once the monkey is trained, the virtual cursor can be unhooked from the monkey's wrist and be guided directly by signals from the animal's neurons. This allows the monkey to observe how its 'thoughts' influence the cursor's movement, and to adjust its responses accordingly. Future users of prosthetic devices will probably use visual feedback in a similar way.

Pump down the volume

Despite the impact of multi-unit experiments on prosthetics and fundamental neuroscience research, some in the field believe that simultaneous recordings from more neurons than the current best of 100 will be needed if the true potential of the technique is to be realized. Just as a computer image becomes sharper with more pixels, the picture of what a brain network is doing should become crisper if the activity of more neurons can be measured. With this in mind, Bruce McNaughton of the University of Arizona at Tucson is now developing techniques to record from several hundred neurons at once.

But for some applications, it is not just a question of sheer numbers, says György Buzsáki of Rutgers University in Newark, New Jersey. The difficulty, he explains, lies in recording from a large number of neurons within a sufficiently small region of the brain. "The big deal is to record many neurons from a very small volume."

Most information processing in the brain is thought to occur within clusters of a few hundreds or thousands of neighbouring cells, which occupy volumes smaller than a cubic millimetre. To understand how these circuits work, Buzsáki thinks it will be necessary to sample the activities of multiple cells within them simultaneously. But the microwires typically used as electrodes cannot be used to record from large numbers of cells in such a small volume, because of the damage their blunt tips would inflict on brain tissue.

The key to getting around this, says Buzsáki, lies in the technology used to build computer chips. Using micromachining and lithography, his team has created slender silicon shafts with up to 16 iridium electrodes

positioned along each shaft's length which can be used to record from different neurons independently¹². In addition to increasing the number of electrodes per implantation site, says Buzsáki, his sharp shafts tend to move through brain tissue without crushing it. His team's latest success is a probe sporting six shafts, yielding a total of 96 electrodes, which span less than a cubic millimetre of brain tissue. And Buzsáki thinks he can do even better. "The currently available industrial technology can produce probes with the same dimensions, but four to five times more recording sites," he says.

Information overload

But the toughest problem for multi-unit studies lies in the analysis of the acquired data. "To my mind, this is the stumbling block," says Moshe Abeles of the Hebrew University in Jerusalem. A single hour of recording from a few dozen cells can produce gigabytes of data, from which researchers must try to extract meaningful information. Many researchers analyse networks of neurons by comparing the activity of each cell to each of the others, one pair at a time. Although these studies provide some clues about how the network functions, they are still far from revealing how tens, hundreds or thousands of neurons behave as a group.

Buzsáki believes the key will be algorithms that can analyse network function by considering many component cells in one go. "Extracting this hidden information requires mathematical tools that are not available anywhere—not in industry, not in academic mathematics departments, nowhere," says Buzsáki. Researchers such as Abeles are now forming multidisciplinary teams, including computer scientists, physicists and mathematicians, to tackle the problem.

If they succeed, neuroscience would be propelled to new heights. But even without a leap forward in data analysis, enthusiasts are convinced that multi-unit recordings are here to stay. "This method is going to be the next revolution in neuroscience, and it's happening right now," says Nicolelis. ■

Marina Chicurel is a writer in Santa Cruz, California.

1. Moore, G. P., Segundo, J. P., Perkel, D. H. & Levitan, H. *Biophys. J.* **10**, 876–900 (1970).
2. von der Malsburg, C. *The Correlation Theory of Brain Function* Internal Report 81-2 (Max Planck Institute for Biophysical Chemistry, 1981); online at: <http://www.neuroinformatik.ruhr-uni-bochum.de/ini/VDM/PUB-LIST/1981>
3. Gray, C. M., König, P., Engel, A. K. & Singer, W. *Nature* **338**, 334–337 (1989).
4. Engel, A. K. & Singer, W. *Trends Cogn. Sci.* **5**, 16–25 (2001).
5. Castelo-Branco, M., Goebel, R., Neuenschwander, S. & Singer, W. *Nature* **405**, 685–689 (2000).
6. Alonso, J.-M., Usrey, W. M. & Reid, R. C. *J. Neurosci.* **21**, 4002–4015 (2001).
7. Louie, K. & Wilson, M. A. *Neuron* **29**, 145–156 (2001).
8. Humphrey, D. R., Schmidt, E. M. & Thompson, W. D. *Science* **170**, 758–762 (1970).
9. Georgopoulos, A. P., Schwartz, A. B. & Kettner, R. E. *Science* **233**, 1416–1419 (1986).
10. Schwartz, A. B. *Science* **265**, 540–542 (1994).
11. Weissberg, J. et al. *Nature* **408**, 361–365 (2000).
12. Bragin, A. et al. *J. Neurosci. Methods* **98**, 77–82 (2000).



Thinking small: György Buzsáki wants to gather multi-unit data from tiny volumes of brain tissue.

Making sure that the world's palaeodata do not get buried

International efforts are keeping information accessible.

Sir — Your Opinion article “Make the most of palaeodata” (*Nature* **411**, 1; 2001) called for public funding of palaeoclimatic research, more interdisciplinary collaboration and the systematic archiving of data.

The International Geosphere–Biosphere programme's core project on past global changes (PAGES) (<http://www.pages-igbp.org>) and the World Data Center for Paleoclimatology (<http://www.ngdc.noaa.gov/paleo>) exist specifically to address these needs, including facilitating interdisciplinary collaboration, providing publicly available palaeoenvironmental data archives, and coordinating the efforts of the “maze of publicly supported databases”.

For more than a decade, the World Data Center for Paleoclimatology (WDC-Paleo), hosted by the National Oceanic and Atmospheric Administration (NOAA) Paleoclimatology Program in Boulder, Colorado, has been working to ensure that palaeoclimate data are freely available. Serving as the PAGES international palaeodata system, WDC-Paleo houses myriad proxy data from thousands of sites. PAGES and WDC-Paleo collaborate with the research community and international partners to provide a long-term, open archive of these important data.

Many agencies and journals are already helping to promote data sharing. In the United States, the National Science Foundation (NSF), the NOAA and the US Global Change Research Program have strong data-sharing policies. The NSF/NOAA initiative on Earth system history requires as a condition of funding that data are submitted to WDC-Paleo. The journals *Paleoceanography* and *Journal of Paleolimnology* both encourage authors to submit their data to WDC-Paleo, and *Science* now requires archival data to be deposited in a publicly accessible database. We encourage more journals to require or request submission of data to internationally approved public archives.

PAGES, WDC-Paleo and the many scientists and institutions that support them have made great efforts to make data easily accessible and usable.

Agencies funding palaeoclimate research must be committed to providing resources for investigators to prepare their data for inclusion, and for data centres to

maintain free, open, international archives of palaeoclimate data and the tools to access them.

Keith Alverson*, C. Mark Eakin†

*PAGES International Project Office, Bärenplatz 2, Bern CH-3011, Switzerland

†NOAA Paleoclimatology Program and World Data Center for Paleoclimatology, NOAA/National Geophysical Data Center, 325 Broadway E/GC, Boulder, Colorado 80305-3328, USA

Who will organize sharing of biodiversity data?

Sir — I welcome your Opinion article (*Nature* **411**, 1; 2001) drawing attention to the issue of how to share the exploding palaeodata relevant to global change. Few of us have begun to face up to the methodological, legal and cultural implications, let alone thought about how to process the data for greatest value.

The issues go far wider than those covered in your Opinion article, leading from palaeodata to information from palaeontology and biodiversity of modern organisms. Data from all these sources need to be assessed and integrated before combined analyses in the interdisciplinary approaches that will be required to understand the latest environmental challenges.

Recent attempts to build multi-query

information-technology systems for biodiversity data are struggling for money and an agreed methodology.

Various databases are separately available on the Internet, covering palaeodata (<http://www-odp.tamu.edu>); the fossil record, through the International Organisation of Palaeobotany (<http://ibs.uel.ac.uk/palaeo>); modern species, through the International Organization for Plant Information (<http://iopi.csu.edu.au/iopi>) and the Man and the Biosphere project (<http://ice.ucdavis.edu/mab>); and quaternary ecosystems through the World Data Center for Paleoclimatology's global pollen database (<http://www.ngdc.noaa.gov/paleo/ftp-pollen.html>). But these databases have resisted the integration that is technically possible.

An example of a multi-query system is now available at <http://biodiversity.org.uk>, where it is possible to query more than 20 databases at different locations (click on ‘search data’ and then ‘Java-enabled version’). For example, if one uses the wild card * and searches for Sal*, one will receive details of salmon, *Salmonella*, *Salix*, salt, salamander and so on from databases situated in the Philippines, Japan, Australia, London and California, respectively.

The OECD is taking responsibility for some of the biodiversity data. It is time to expand the support to include these other fast-growing data as well.

One central multidisciplinary agency, such as the International Council for Science or UNESCO, must be in control. Who is to make the nomination?

Michael Boulter

Palaeobiology Research Unit, University of East London, Romford Road, London E15 4LZ, UK

Mistaken identity in a watery arms race

Sir — The thought-provoking Concepts essay “A watery arms race” by Victor Smetacek (*Nature* **411**, 745; 2001) concerns defence systems in the phytoplankton, yet the illustration shows a testate amoeba — neither planktonic nor photosynthetic!

I like to think that most students of freshwater biology 20 years, even 10 years, ago would have recognized the incongruity. But what leaves me frightened for the future of whole-organism biology is that *Nature*, the prestige biology journal, should be guilty of such a howler.

Doubtless the picture was sought in a

hurry and in the interests of immediacy. This is all very well for those who set store by citation indices, but the consequence is that the article may be mentioned as much for the wrong reasons (as an example of the demise of whole-organism biology) as for the right ones. It is scarcely fair to the author, to his interesting ideas or to our discipline.

Richard M. Crawford

Friedrich Hustedt Diatom Collection, Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany

This error was *Nature's*, not the author's. The caption supplied by the photo library was incorrect — Correspondence Editor, *Nature*.

Mission now possible for AIDS fund

Adequate support by the G8 countries is needed to defeat this global killer.

Peter Hale and colleagues

The G8 summit in Genoa this week (20–22 July) may be the last chance to increase funding for the global fund on AIDS on the scale required — and to get it up and running this year. Waiting another year will cost countless lives. Donors and recipient countries now agree that the fund must be sustainably financed, properly structured, have a clear mandate, be lean and agile, and be independent to decide on the distribution of funds. But consensus is meaningless without adequate resources.

Kofi Annan, the United Nations (UN) secretary-general, has said that US\$7 billion–10 billion a year is needed for a comprehensive programme for HIV/AIDS prevention and treatment in the poor and middle-income countries most affected by the epidemic. Two independent cost estimates published this year^{1,2} have produced similar figures. The UN team² estimates that \$9.2 billion per year will be needed by 2005, in line with confidential government and industry assessments, as well as other private estimates currently in preparation. Careful comparison of these estimates suggests that \$8 billion–10 billion per year will be needed over the next 5–10 years, not all through the AIDS fund. Recipient countries are expected to cover a substantial portion of this amount annually. But for the least-developed countries in Africa, and south and southeast Asia, at least 80% of the funds will have to come from international sources².

Yet, despite public and political support in the West, commitments to the Global AIDS Fund total less than \$1 billion (see Table 1). This is not enough. Donor coun-

Table 1 Commitments to the Global AIDS Fund to 10 July 2001

Country	Commitment (US\$ million)
United States	200
United Kingdom	200
Japan	200
France	127 over 3 years
Gates Foundation	100
Nigeria	10
Luxembourg	2.5
Uganda	2
Zimbabwe	1
Austria	1
Winterthur	1
Total	844.5

Only commitments over US\$1 million are listed — there are smaller contributions, which add up to less than \$500,000. Canada and Italy are expected to announce their contributions at the G8 summit; Germany and Russia have yet to hint at theirs. Source: UN Foundation.

tries are fooling themselves if they feel that this would make a significant impact; it will not. It is also unclear whether the \$845 million pledged so far will be new money or whether it would be diverted from other development programmes. Of the \$8 billion–10 billion required to endow the fund each year, roughly half will be for prevention efforts in highly affected countries; the other half will be for improvements in infrastructure to expand access to treatment².

Although \$8 billion–10 billion is a lot of money, it is in line with AIDS spending elsewhere. For example, the United States has more than 800,000 people living with HIV/AIDS and spends \$20 billion a year on prevention and treatment. Countries in the European Union spend only slightly less per head. In this context, \$8 billion–10 billion for

33 million people living with HIV/AIDS in the least-developed nations in Africa, Asia, Central and South America is not unreasonable. Moreover, this amount represents just 0.005% of the total gross national product of the seven richest G8 nations. Last month, Donald Rumsfeld, the US secretary of defence, announced plans to order 60 B2 bombers at a total cost of \$120 billion. This sort of cost comparison is often criticized, but if we agree with US secretary of state Colin Powell that fighting AIDS is like waging war — and we do — it is reasonable to compare the costs of what it is going to take to beat the enemy. In this case, the enemy has already killed 23 million and will kill at least 3 million more this year.

There has always been a disconnection when it comes to financing the war on AIDS. To bring the total up to \$8 billion–10 billion, each donor country needs to add a zero. A one-log increase by each of the G8 nations will put the fund in reach of this goal. The extra needed, \$1 billion–2 billion, should follow from industry and the private sector, especially if tax incentives are provided.

If we don't act now, AIDS will cost billions of dollars more later. Instead of adding one zero now, we may have to add two zeros in ten years — and witness during this time the horrific human and socio-economic toll of an epidemic that will decimate Africa, and perhaps parts of the Caribbean, within a decade. Failure to act now will also allow the virus to continue its rampage across Asia and Eastern Europe, and will impede efforts to control its spread in Latin America.

For far too long, many African and Asian leaders have wasted precious time in failing

HIV a '9' on Richter scale of viral diseases

HIV is, without doubt, the 'big one'. This virological 'Richter' scale ranks viruses in the same way as earthquakes. The figure is based on a log₁₀ scale, where 10 represents all deaths worldwide from infectious diseases in 1999. HIV registers a 9.1. For HIV, 23 million people have already died, 36 million are incubating the virus, and 5 million are

newly infected each year. More than 3 million will die this year, mostly in sub-Saharan Africa.

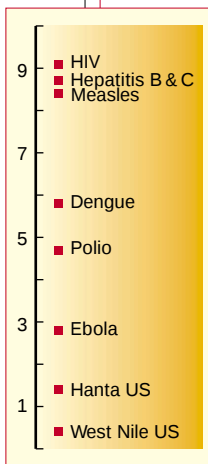
The only viruses that have come close to this doomsday scale are the Spanish influenza that killed more than 20 million in the 1918–19 pandemic, and smallpox, which has now been eradicated. Today, hepatitis B and C are major causes of liver disease and cancer, collectively ranking 8.7 in terms of deaths. An effective vaccine against hepatitis B is being introduced worldwide.

Several orders of magnitude lower are dengue fever and polio. In the 1950s, polio caused a great scare

in the West. An endemic virus, it might have ranked 6.5 at the time. Large-scale vaccination programmes in its last few regional hold-outs, mainly central Africa and India, mean that the conquest of poliovirus is within reach. Ebola ranks a 3 at most; although frightening, it does not seem to travel well. The West Nile and Hanta viruses, both of which made front-page news in the United States, barely cause a ripple on the scale. A death from Ebola is as great a loss as one from HIV, but in terms of public health it is madness to lose sight of the big ones.

Of all infectious diseases, HIV is the only one that has skyrocketed

from nowhere to more than 60 million infections in around 30 years. Given its sexual transmission among young adults, its highest prevalence among the poorest nations of the world, the cost of antiviral medicines, and the lack of a vaccine, there is no let-up in sight. There is nothing to suggest that HIV will plateau, or that it will not reach one billion cases before 2050. It is unlikely, but not impossible, that HIV could attenuate itself, but when and at what human toll? Nobody has any idea. Intervention must be on the same massive scale as the magnitude of the epidemic.



to confront the epidemic head-on. Now that a growing number of them are facing the issue, we cannot afford Western leaders to go into denial about how much it is going to cost to fix the problem. For the first time, the world is united in spirit behind an ambitious plan to curb the spread of HIV. The unprecedented declaration accepted by 185 states at last month's special UN General Assembly session on AIDS is compelling evidence of a new willingness by the hardest-hit countries in Africa and Asia to step up their prevention programmes and to make the necessary improvements in health infrastructure.

We, the authors of this Commentary, collectively have more than 150 years' experience of HIV/AIDS, as scientists, clinicians or public-health experts. We believe that the Global AIDS Fund, if properly financed and

managed, represents our best chance to stem the epidemic. How the fund is managed and run needs to be determined, but it must not become a turf war between development agencies and stakeholders. There is room for everyone's ideas to be included.

Decisive action by the G8 nations is crucial in determining whether the fund is successfully launched this year on the scale required. The success of the rich, industrialized nations is inextricably linked to the success of the developing nations. This is why we hope that the G8 countries will rise to the occasion and find ways to finance the fund to accomplish its mission. Such action will not only be a mark of true leadership, but also of humanity in its highest form. ■

1. Sachs, J. D. *Nature Med.* **7**, 521–523 (2001).

2. Schwartländer, B. et al. *Science* **292**, 2434–2436 (2001).

Success hinges on support for treatment

HIV prevention and treatment are inseparable. One big debate that never materialized during the UN General Assembly meeting on AIDS last month (see *Nature* **411**, 984; 2001) was about priorities for the Global AIDS Fund, particularly how to allocate funds to prevention compared with those for improvements to public-health infrastructure and access to treatment with anti-HIV drugs. Most AIDS experts endorse the idea that prevention and treatment are crucially linked, and the authors of a policy forum in *Science* (**292**, 2434–2436; 2001) estimate that the split should be roughly 50/50. Unfortunately, not all government and private development agencies agree, nor do some high-ranking Western officials.

The overwhelming majority of Africans, of course, want treatment; the question is not 'if' but 'when'. "Don't even ask us," said one African delegate. "The answer is yes, yes, yes and yes." You can ask the same question of any young man or woman on the streets of Soweto or Lusaka and get the same answer.

Any debate on prevention versus treatment would not have been possible even a year ago because of the prohibitively high cost of combination anti-retroviral therapy. Since then, drug companies have discounted the cost of these medicines for the least-developed countries, sometimes by as much as 90%. So far, 58 nations have purchased HIV/AIDS drugs at preferential prices, bringing the cost of treating HIV closer to that of treating chronic conditions such as type-2 diabetes and high blood pressure.

Prevention and care are synergistic. Attempts to prioritize one at the expense of

There is little incentive to get tested for HIV if there is no treatment.

the other are morally indefensible, a denial of a fundamental human right, and just plain bad public health. The main argument for focusing on prevention rather than treatment is that it is more cost-effective when funds are limited. This masks the mistaken but still widely held view in the West that treatment in poor countries cannot be funded, even with discounted drug prices, because of the lack of basic health-care infrastructure (trained doctors and nurses, hospitals, clinics, labs and equipment).

Yet considerable infrastructure exists in countries such as South Africa, Kenya and Zimbabwe. Where there is political will, infrastructure can be upgraded on a crash basis. Human ingenuity to create temporary structures to do the job effectively should not be underestimated; many Western hospitals boast trailers and temporary buildings yet deliver world-class medical care. Certainly, money is needed for infrastructure, but a little goes a long way in Africa.

Another myth is that Africans will be unable to follow complex drug regimens, leading to the development of resistant virus that could be transmitted. On the contrary, studies in Africa, especially in Uganda and Senegal, show that compliance with drug

regimens where there is patient education is as good as in New York City. In any event, regimens are nowadays much more simple.

A third myth is that the standard of HIV care would be suboptimal, so it should not be attempted. This hypocritical view overlooks the beginnings of HIV treatment in the West (monotherapy, then bi-therapy, then triple therapy), as doctors and patients learned as they went along. It was distressing to hear this argument advanced by a few African officials after the UN meeting. It also sets an impossibly high standard for expanded access to HIV care for the vast majority of Africans who are poor, unemployed or without health insurance.

Finally, treatment with anti-retroviral drugs helps prevention efforts. There is little incentive for people to get tested for HIV if there is no treatment. An HIV-negative result is a prime opportunity to deliver prevention messages; for a positive test the prospect of treatment increases awareness, removes stigma and encourages safe practices — all of which reduce the rate of new HIV infections.

There is now unstoppable momentum to address the challenge of how to expand access to HIV care and treatment in low- and middle-income countries. For the least-developed nations, including all of sub-Saharan Africa, heavily discounted drugs are available. For middle-income countries, such as Brazil, continued local manufacture of anti-HIV drugs or importing of generic versions is to be allowed until the crisis is controlled.

The UN meeting was intended to intensify national and international action, and to mobilize the billions of dollars needed to combat the epidemic. It was successful in the first respect, particularly in terms of commitments made to specific prevention targets. But the breakthrough was the agreement that the Global AIDS Fund should also cover treatment. We hope that the G8 leaders will respond not just with more money but by mandating the fund to tackle treatment. ■

Corresponding author: Peter Hale, Institut Necker, Faculté de Médecine Necker-Enfants Malades, 156 Rue de Vaugirard, 75015 Paris, France.

Co-authors: Malegapuru William Makgoba (President, Medical Research Council of South Africa); Michael H. Merson (Professor and Dean of Public Health, Yale University School of Medicine); Thomas C. Quinn (Professor of Medicine and International Health, Johns Hopkins University, Baltimore); Douglas D. Richman (University of California, San Diego); Stefano Vella (President, International AIDS Society, Istituto Superiore di Sanità, Rome); Fred Wabwire-Mangen (Director, Institute of Public Health, Makerere University, Uganda); Simon Wain-Hobson (Institut Pasteur, Paris); Robin A. Weiss (University College London).

The authors thank Helene Gayle (CDC, Atlanta) for discussions and suggestions, and Bernard Schwartländer (UNAIDS) and R. P. Eddy (UN Foundation) for fact-checking.

China on the losing side

Mao's drive to conquer nature had shocking consequences for his country.

Mao's War Against Nature: Politics and the Environment in Revolutionary China

by Judith Shapiro

Cambridge University Press: 2001. 306 pp.

£35, \$59.95 (hbk), £12.95, \$18.95 (pbk)

Crispin Tickell

Between 1949 and 1976, Mao Zedong enjoyed supreme political power in China, and applied what he called Marxism–Leninism–Mao Zedong thought to the environment and the people. One of his most potent slogans was “Man must conquer Nature”, and throughout the land, to varying degrees, a kind of conquest took place. But as the economist Fritz Schumacher once wrote, “man talks of a battle with Nature, forgetting that if he won the battle, he would find himself on the losing side”. In this case, China was certainly the loser.

There are three broad philosophical traditions in China. One springs from Taoism, which seeks accommodation with nature and is best encapsulated in the slogan ‘harmony between the heavens and human-kind’; a second springs from Buddhism and its reverence for all living creatures; and the third springs from Confucianism with its respect for authority, including control and mastery of the natural environment. Although Mao was most influenced by the Soviet experience — its forced industrialization, collectivization of agriculture and overturning of established hierarchies — he falls well within the Confucian tradition and relied upon it crucially when applying his ideas and policies.

Judith Shapiro has had a long experience of China. She was a teacher there in the period following Mao's death, and has returned from time to time, most recently last year. Her book is the product of discussions with many kinds of Chinese people, including some of the surviving victims of the Mao years. Her thesis has three main aspects.

The first relates to Mao's policy framework. Like others at the time, Mao associated the recovery of China as a world power with an increase in human population. Then, whatever might happen in terms of war or other disasters, the Chinese people would survive. Families were given every encouragement to reproduce. With greater stability in China, numbers rose dramatically from around 600 million in 1923, and 1 billion when Mao died in 1976, to about 1.3 billion today.

As in the Soviet Union, industrialization was pursued at all levels. Every town and village was encouraged to have its own industrial activity. More drastic still was the



A cultural call to arms: Mao's conquest of nature was treated as a military campaign.



development of agriculture, in particular the production of grain. Mountains were to be moved, forests cut, dams built, lakes filled up, rivers diverted, wastelands planted and errant species eliminated (hence the slogan “Wipe out the four pests” — these were sparrows, rats, flies and mosquitoes). A central example was given for universal application: no matter that this model — the so-called Dazhai model — was in many ways fraudulent.

Shapiro also looks at the effects on the environment. They were soon evident, but Mao would not change course. The Chinese environment has long been fragile and vulnerable, and the policies of the 1950s made it more so. But Mao believed that what had worked for other industrial countries should also work for China. He ignored the consequences of over-extraction of

resources, pollution of air, land and water, erosion of hillsides, flooding and salinization of irrigated areas, reclamation that led to desertification, and all the other ills of unsustainable human activity. A 1994 paper by Qu Geping, one of China's leading environmentalists, and Li Jinchang produced a telling statistic: “from 1957 to 1980 the annual net loss of cultivated land averaged 545,000 hectares.”

Finally come the effects on the people. Mao's programme allowed no argument or opposition. It was ruthlessly directed from the centre with little or no regard for local or even regional interpretation. It involved repression not only of other leaders of all kinds, including scientists, but also of peasant communities. People were forcibly resettled and communities disrupted. The conquest of nature was treated as a military campaign, where leaders were given army rank and powers of command.

Two phases will survive in Chinese history: the Great Leap Forward, which led to one of the greatest of human famines, with deaths calculated at between 36 and 50 million from 1959 to 1961; and the Cultural Revolution, which began in 1966 and overturned Chinese society, causing the loss of a whole generation of leaders.

Mao's attempt to conquer nature is an extreme example of what has been tried by other societies at other times and under other ideologies. The changes he brought about in a society ravaged by civil war but deeply conservative in character may not all have been bad, and some of them, driven by population pressures, might have happened anyway. But it is still a model of what not to do.

BETTMANN/CORBIS

BETTMANN/CORBIS

Subsequent Chinese governments have been trying to pick up the pieces, and put rational policies together again. In doing so they have had to cope with public cynicism, and with vested interests that rely on inertia and the mythology that survives from the Mao years.

There is now another danger, for the shift from a centralized to a market economy could all too easily increase the environmental damage. But this has produced an opportunity, so far barely recognized elsewhere, to establish true environmental costs and thus regulate the economy in the public interest. In 1992 the Chinese government created a China Council for International Cooperation in Environment and Development. It consisted of both Chinese and foreign participants, and had direct access to the leadership. It is already having effects on the formulation of a policy that might this time make China the pioneer.

Shapiro's well-written book could be misinterpreted as a polemic. It tells a shocking story that needs to be told, but ends on a note of hope. The Chinese people now have the chance to re-establish the traditional harmony with nature that was once their ideal. They deserve to succeed. ■

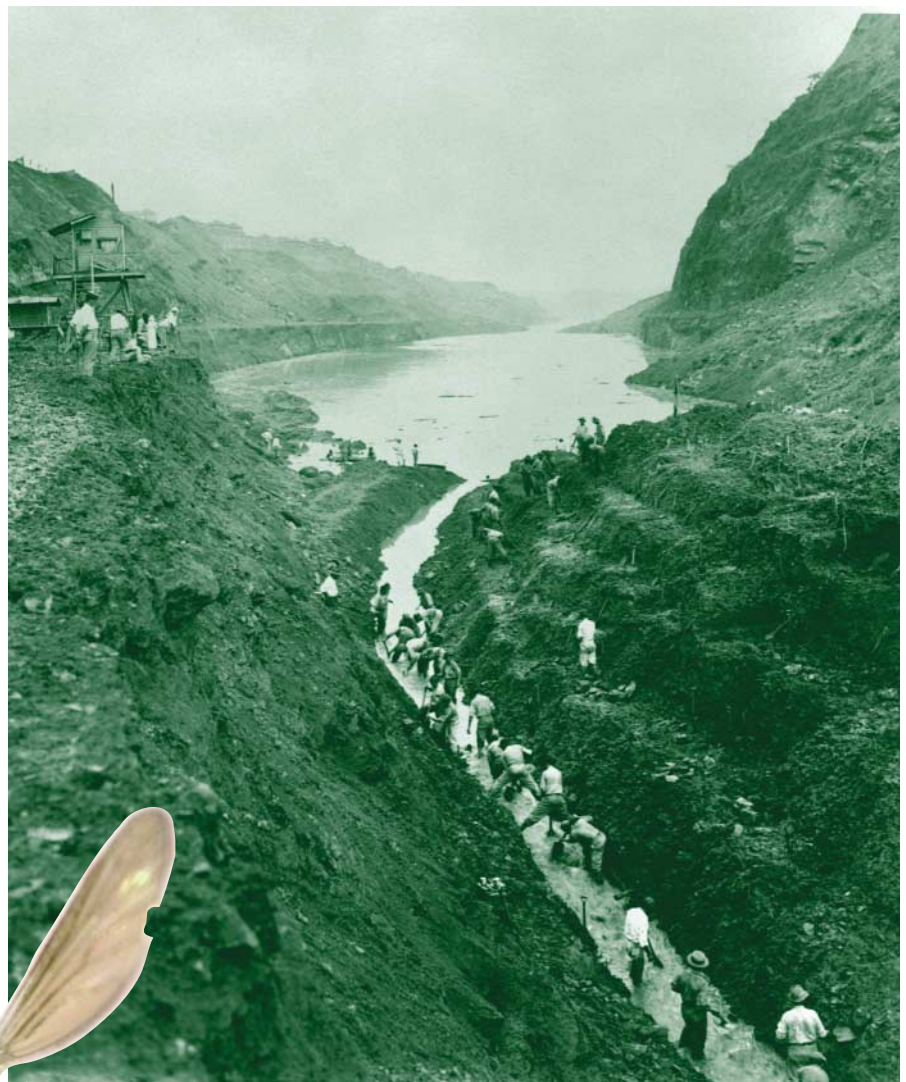
Crispin Tickell is at Abington Old Barn, Abington, Cirencester GL7 5NU, UK.

The battle against lethal vectors

**Mosquito: The Story of Mankind's Deadliest
Foe/Mosquito: A
Natural History of
Our Most
Persistent**

and Deadly Foe
by Andrew Spielman
& Michael D'Antonio
*Faber & Faber/Hyperion: 2001.
247/267 pp. £10.99/\$22.95*
C. F. Curtis

This is a very personal account of mosquito-borne disease by Andrew Spielman, a respected mosquito biologist, assisted by journalist Michael D'Antonio. The emphasis is strongly on the Americas, so that considerably more space is devoted to the recent outbreak of West Nile fever in New York than to the incomparably more serious problem of malaria in Africa. Apart from an interesting account of the physician Patrick Manson's discovery in the 1880s in China that filarial parasites are transmitted from infected humans to mosquitoes, almost nothing is said about filariasis. And yet this mosquito-borne disease is considered to be the second-largest cause of disablement worldwide after mental illness.



Making the link: the Panama Canal was eventually completed once the importance of mosquito control in combating the yellow fever and malaria that decimated workers was recognized.



The book contains interesting accounts of the many aspects of mosquito biology to which Spielman has contributed. However, a medical entomologist requiring a thorough account of any particular aspect of these subjects would be better advised to go to Alan Clements' multi-volume *The Biology of Mosquitoes* (Chapman & Hall, 1992, 1999).

The real strength of Spielman and D'Antonio's book is its fascinating coverage of the important historical events related to mosquito-borne diseases and their control in the Americas. The crucial importance of yellow fever in earlier centuries is underlined by an account of the devastating outbreak in Philadelphia in the 1780s, at that time the capital city of the United States. The fear that yellow fever would continue to reach the United States from Cuba was one of the factors in the murky history behind the outbreak of the Spanish-American War at

the end of the nineteenth century. This was the era in which one tropical disease after another was being shown to be transmitted by insect vectors.

The book gives a careful account of the proof that urban yellow fever is transmitted by the mosquito *Aedes aegypti*. In the days before ethical committees, proof was gained by allowing mosquitoes to feed first on patients and then on healthy subjects; the 'controls' were healthy subjects who slept swathed in sheets soiled with the 'black vomit' characteristic of yellow fever. Recognition of the vital importance of mosquito control in combating yellow fever and malaria led to the completion of the Panama Canal by the United States in 1914, after French attempts 20 years earlier had been defeated by these diseases.

The Rockefeller Foundation's attempt to extend the success in Panama to the whole of Brazil meant that there were mosquito experts there who could identify the world's most efficient malaria vector, *Anopheles gambiae*, when it was accidentally intro-

duced into Brazil from Africa in the early 1930s. The subsequent malaria epidemics and the eventually successful eradication of *A. gambiae* from Brazil after a ten-year struggle, directed by the formidable Fred Soper, are graphically described. These events in Brazil (and parallel events in Egypt that are not mentioned), together with the arrival of DDT and the desire of the US authorities to demonstrate the benevolent power of their technology, led to the attempt to eradicate malaria worldwide.

This is now sometimes held up as an absurd example of over-confidence in technology. But one should not forget the historical context and how successful the campaign was in many areas, so long as the political will was maintained to keep the spray campaigns going. Spielman and D'Antonio describe the reaction against DDT following the publication of Rachel Carson's *Silent Spring* in 1962. And they bring the DDT story up to date with the conclusion of the treaty on Persistent Organic Pollutants in December 2000, which exempted DDT from being banned because of its still-vital role in vector control.

Spielman has long opposed the sometimes glib proposals that the malaria problem can be solved by driving genes into wild mosquito populations that can prevent the insects from becoming infected with the *Plasmodium* malaria parasite. He emphasizes the real difficulties in securing an unbreakable linkage of the desired gene to the driving system that is intended to introduce this gene into continent-wide mosquito populations. He also points out (with some personal reminiscences) the problems of persuading awkward individuals to agree to the community-wide application of anti-mosquito measures (especially if transgenic mosquitoes are to be released). But the authors seriously misstate the reasons for the ending of a research project near Delhi in India in the 1970s to test the feasibility of controlling mosquitoes by releasing sterile males. The problem was not, as they state, that villagers believed they themselves would be sterilized, but stemmed from the fantasies of a journalist and a few politicians that the project was really a sinister US plot to study the use of *A. aegypti* and yellow fever for biological warfare against India.

In a rather brief final section on possible future control methods, Spielman describes his own recent work showing that pollen from local maize crops is an important food source for *A. gambiae* larvae. He makes the exciting suggestion that transgenic maize, carrying a gene for a toxin found in *Bacillus thuringiensis israelensis*, could be a way for farmers to produce a potent and specific larvicide that would be effortlessly distributed into breeding sites by wind-borne pollen. ■
C. F. Curtis is at the London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK.

Transgenic ills and otherwise

Redesigning Life? The Worldwide Challenge to Genetic Engineering
edited by Brian Tokar

Zed Books: 2001. 440 pp. \$19.95, £15.95

Bernard Dixon

Just as good government depends on effective opposition, so new science benefits from peer review, and novel applications of science need their technical and social critics. There is reason, then, to welcome a work described as "the first comprehensive examination of the hidden hazards of the new genetic technologies and the emergence of worldwide resistance".

Edited by Brian Tokar, an opponent of genetic modification (GM), *Redesigning Life?* contains 31 chapters by other GM activists, members of non-governmental organizations and journalists. An opening section on health, food and the environment is followed by another covering medical genetics, science and human rights. These are complemented by sections on opposition to genetic engineering and on "patents, corporate power and the theft of knowledge and resources".

Whether sympathetic or otherwise to this cause, readers will learn much here about the foundations of hostility towards recombinant DNA manipulations, and about campaigners' tactics. However, those seeking balanced or simply fair-minded discussion will be deeply disappointed.

For example, several contributors discuss the "scandal" of bovine spongiform encephalopathy in Britain and the related emergence of one form of Creutzfeldt-Jakob disease in humans. Yet there is no mention of the fact that human growth hormone, formerly extracted from the pituitaries of human cadavers, can now be produced in recombinant bacteria, thereby eliminating the tragedy of avoidable Creutzfeldt-Jakob disease in patients who need growth-hormone treatment. This "comprehensive" book is similarly silent on genetically engineered erythropoietin, interferons, blood-clotting factors and a wide range of other recombinant products that have been used safely and to the lasting benefit of patients for many years.

The approach of Tokar and his colleagues is illustrated by three items that do appear in the book. One stems from a 1999 article in *The Guardian* newspaper claiming that the British Diabetic Association had withheld evidence that thousands of diabetics may have suffered a deterioration in health after using recombinant insulin. Although the association has categorically denied both

the alleged evidence and its alleged cover-up, the charge surfaces again here in a piece about "a suppressed 1993 report commissioned by the British Diabetics [sic] Association".

Also questionable are the sources given to support the claim by researcher Árpád Pusztai that potatoes genetically modified to carry a lectin gene produced adverse effects in rats. These claims triggered the GM food furore in Britain in 1998. We are told, for example, that the amounts of protein, starch and sugars in the tubers differed by as much as 20% from non-GM potatoes. However, the reference is not to a peer-reviewed journal but to an interview in something called *GM-Free*.

Similarly, when rats ate the GM potatoes, "the immune response was significantly depressed and there was evidence of serious intestinal inflammation and infection". The reference here is to Pusztai's 1999 *Lancet* paper, which (though greeted in the media as vindicating his original claims of adverse effects on growth and the immune system) contains no evidence on these aspects, nor on infection. The paper is about the structure of the small intestine.

Third, we read that "a single intentional change in a plant's DNA can cause other, unintended health and safety problems". Yet the single citation here is to a 1996 *New England Journal of Medicine* paper about an allergen gene from brazil nuts. The authors of the paper showed that this can be expressed in soybeans, which provoked allergic reactions in people already known to be allergic to brazil nuts. So what?

The real purpose of that research was to investigate and then remove a potential hazard. Alas, Tokar and his colleagues do not seem interested in these positive aspects of transgenic work. They do not even mention the most exciting potential of genetic manipulation regarding allergy — the possibility of eliminating potential allergens from foodstuffs and thus avoiding the misery, and worse, that they can cause.

There are, of course, genuine environmental, social and geopolitical grounds for caution over the robust burgeoning of biotechnology, and some of them are explored in this book. Unfortunately, however, its overall tone is summarized all too clearly in the words of one contributor, Zoe Meleo-Erwin, who writes: "Reproductive and genetic technologies are not designed to help the infertile, cure the sick, improve upon evolution or offer greater choice. This industry seeks power over nature, women, children and human evolution in the pursuit of increasing profits."

Many references in *Redesigning Life?* are to newspapers and magazines, reminding us of the role the media have played, particularly in fomenting hysteria over GM foods. That influence is examined crisply and

perceptively in *A Grain of Truth: The Media, the Public, and Biotechnology* (Rowman & Littlefield, 2001) by Susanna Hornig Priest, a professor of journalism in the United States. It should be read carefully by all who believe that the media get everything wrong, that journalists exercise excessive power over a gullible public or that they are too preoccupied with either doom or hype. Reflective reading of this book might help to bring journalism, biotechnology and even its opponents into a more fruitful relationship. ■

Bernard Dixon is at 130 Cornwall Road, Ruislip Manor, Middlesex HA4 6AW, UK.

Not quite the whole story

The Story of Mathematics

by Richard Mankiewicz

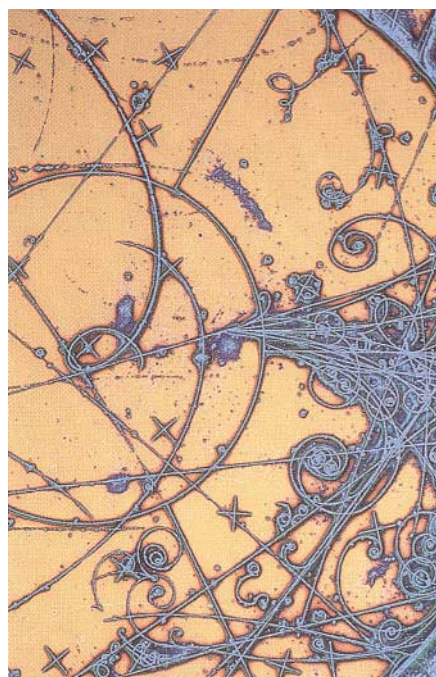
Princeton University Press/Cassell: 2001.
192 pp. \$24.95/£18.99

Karine Chemla

The Story of Mathematics guides the reader from prehistoric times to the era of 'chaos and complexity' in 191 lavishly illustrated pages, many of which are magnificent reproductions of mathematical documents, instruments or visual evidence of mathematicians' work. No wonder that, as the preface concedes, ruthless selection was needed in the material covered. Indeed, the author has devoted some attention to the mathematics of most periods, and every ancient tradition in particular. But the severe selection has had a notable effect on the range of topics covered. Mathematics in the twentieth century, for example, has mainly been boiled down to four chapters dealing, respectively, with games, modern art, machines and chaos. So what principles guided the choice of the "key developments in mathematical ideas" in given historical periods?

Richard Mankiewicz's main concern seems to be to bridge the gap between mathematics and the general 'culture'. First, by showing how the development of the mathematical sciences in each civilization was related to the historical context, Mankiewicz hopes to contribute to shaping a "scientific culture" today. This aim probably accounts both for the criteria used to select the mathematical topics and the presence of historical description that cannot always be easily related to mathematics. His second strategy to bridge the gap between the 'two cultures' is to show the long-standing interaction between 'artistic sensibility' and mathematics.

Mankiewicz's reasoning seems to be as follows: the beauty of mathematics can be made accessible through its visual aspects.



DE POLYGRAPHIE 176		
Ordres antiques lettres Numerales		
DC	60000000	500000000
DCC	70000000	600000000
DCCC	80000000	700000000
DCCCC	90000000	800000000
XX	100000000	900000000
XXI	110000000	1000000000
XXII	120000000	1100000000
XXIII	130000000	1200000000
XXIV	140000000	1300000000
XXV	150000000	1400000000
XXVI	160000000	1500000000
XXVII	170000000	1600000000
XXVIII	180000000	1700000000
XXIX	190000000	1800000000
XXX	200000000	1900000000
XXXI	210000000	2000000000
XXXII	220000000	2100000000
XXXIII	230000000	2200000000
XXXIV	240000000	2300000000
XXXV	250000000	2400000000
XXXVI	260000000	2500000000
XXXVII	270000000	2600000000
XXXVIII	280000000	2700000000
XXXIX	290000000	2800000000
XL	300000000	2900000000
XLII	320000000	3100000000
XLIII	330000000	3200000000
XLIV	340000000	3300000000
XLV	350000000	3400000000
XLVI	360000000	3500000000
XLVII	370000000	3600000000
XLVIII	380000000	3700000000
XLIX	390000000	3800000000
L	400000000	3900000000

Tay iniques icy deduit, & décrit par forme d'exemple, le moyen & methode d'écrire en lettres Latines & communes, les nombres & la mode numerale des Anciens: à fin que d'eux les ignorants en eussent plus facile intelligence. Et quant à ce, qu'en la suite de description & apposition des figures d'algorithm, j'aurois en quelques endroits laissé l'ordre & règle de la vraye supputation & forme de compter, ce n'a esté par erreur, ny aussi par faute ou omission. Mais par bonne & raisonnable cause j'ay aduisté ainsi le faire: & meismement à fin que par trop grand progression & prolixité ne donnasse ennuy à moy, ny aux lecteurs, qui en la figure subsequeute, briefvement cy apres transcrite, verront autres moyens & ordres numeraux, par lesquels les anciens manifestotent leurs grands nombres.

Integrating maths: these images of a celestial chart (top left), roman and ordinary numerals (right) and particle tracks are aesthetically pleasing, but do they make mathematical ideas more accessible?

"All ideas," Mankiewicz claims, "are born of a vision." Hence the multiplicity of illustrations in the book. One might wonder, however, what mathematical ideas and inner conceptual beauty are conveyed by aesthetically pleasing pages of writings that are reproduced without explanation, and whose artistic qualities remain unconnected to their content. Is this the sense of beauty that does justice to mathematical ideas? Is this the way in which we want mathematics to become an integral part of the culture? Such photographs don't seem the best way of showing how mathematicians use visual auxiliaries, such as figures and models, to work on mathematics and to illustrate or produce their ideas. That would seem to me to be a more adequate use of illustrations to convey ideas, a point I do not find addressed in the book.

On a more technical front, the book contains inaccuracies of fact and interpretation. To mention just two of them, it is not the case that, as Mankiewicz states, in Babylonian mathematics, "arithmetical calculations were handled much as we would today". Moreover, one cannot speak without qualification of the "justifications" that Babylonian mathematicians gave to their procedure for solving quadratic equations. The bibliography, which contains an odd selection of titles, is given without dates or publishers' names. The lack of care given to the text contrasts strangely with the undeniable magnificence of the illustrations. ■

Karine Chemla is at REHSEIS (Recherches Epistémologiques et Historiques sur les Sciences Exactes et les Institutions Scientifiques), CNRS—Université Paris 7, 75005 Paris, France.

Walking with producers

Is science on television often edited for impact rather than accuracy?

Simon Lamb

Imagine the following exchange one evening at your local pub. "I have just seen this amazing programme on the box. These scientists have discovered that dinosaurs once lived on Mars and were transported to Earth by aliens." "Yeah, I really like those science programmes. They are so interesting."

The next morning, on the other side of town, in the coffee room of a university geology department: "I saw a dreadful documentary last night. All that nonsense about the evidence for life on Mars. And Bloggs made a real fool of himself." "What do you expect? It's only TV, after all."

Parody, certainly, but there is a grain of truth, too. Science on television both fascinates and confuses a non-scientific audience. But it is often fundamentally unsatisfying to scientists. Nevertheless, television is probably the main medium through which most people learn about scientific ideas.

Why do scientists seem unable to get their message across? After all, they are interviewed on camera, and they are not knowingly going to talk rubbish. But it is never quite like that. For a start, in my experience, you are usually approached out of the blue for odd pieces of isolated information. You will have little idea how your contribution fits into the overall programme. And then the editing changes everything. Juxtaposition of material can decide how the individual components come across. For example, a recent prime-

time TV documentary, ostensibly about the geology of the planet Venus, intercut comments on planetary physics by a leading geophysicist with the ceremony of a religious cult who seemed, as far as I could tell, to believe that aliens live on Venus. This made the geophysicist look foolish and boring.

It is common practice to cut and even, sometimes, to change the words the scientists actually say — all, of course, for the sake of clarity. How? It is a simple matter to snip out the caveats or preamble that you have carefully put round your scientific statement. The mays, mights, possibly, ifs and buts can be removed. This way a cautious and tentative conclusion becomes a well-known fact. Finally, it is all stitched together with a commentary. This, more than anything else, determines the tone and content of the programme.

I learnt a lot when I approached a major TV science strand to make a documentary on the idea that continents can flow, behaving essentially as a fluid. The response was very cool. I had failed to understand where TV producers are coming from. As journalists, they are after a human story. And as film makers, they want to make something that is visually exciting. So I repackaged the science in terms of the pioneering discoveries of a remarkable New Zealand geologist. I was getting warmer. A human story was beginning to emerge. But it apparently was still not visually interesting enough for television. I tried once again, this time emphasizing that the New Zealander had started his career as a gold prospector, and it was this experience that had ultimately led to his major scientific discoveries. Bingo! I had a visual story with a personality, and the documentary was eventually made.

I was lucky to be involved in all stages of the making of this programme. And I found that this was the only way to ensure that my vision of the science got across. Decisions made at each stage can be very difficult to undo at the next. But, without doubt, it is in the editing and commentary writing that you have the most control.

There is a tremendous temptation to exaggerate. Lines of commentary such as "a startling discovery that has unlocked the mysteries of the Universe, changing for



Smile for the camera: TV producers are after a visual story with a 'personality'.

BBC WILD

ever the way scientists think", or "the most violent event to affect the planet since its creation", are so easy to write. This is just the start of the slippery slope to yet another overblown and sensational science programme. For example, the scientists have not discovered what we hoped they would. It is just not exciting enough. Couldn't we make A's speculative ideas more into fact? How are we going to handle the fact that it was B, not A, who did the research? Perhaps we should just focus on A, or B? Who sounds better on camera? And that bit of scientific evidence is just too complicated for TV — let's skip it. And so on.

In the editing room, I began to feel that I was a sort of quality-control officer. Time and time again the editor or producer would ask: "Couldn't we just say ...?" "No!" "Surely, we could say ...?" "No!" "Well, then, what can we say?" And we would have to roll up our sleeves and thrash it out. If we were faithful to the science, the final result was always better than any amount of fudging.

In the end we managed to explain to my satisfaction the new theory of mountains, including plenty of field geology and fundamental physics — in fact, all the things I had originally been told were boring. And the television executives thought it was fascinating! The whole process was hard work, but I never forgot that the final programme would be watched by millions of people who, I believe, really want to understand how scientists think and work. ■

Simon Lamb is in the Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK.

There is a tremendous temptation to exaggerate. This is just the start of the slippery slope to yet another overblown and sensational programme.

The fog that was not

Ahmed H. Zewail

In macroscopic systems, we can use classical mechanics to describe an object's motion with precise knowledge of where the object is and where it is going. But the microscopic world is governed by the laws of quantum mechanics, which rob us of this precision. We can no longer simultaneously predict the object's position and momentum, or its energy measured over a finite period of time. We accommodate this by recognizing two of the most powerful and yet indigestible concepts: the uncertainty principle and the particle-wave duality of matter.

The duality concept states that all particles, even molecules, can behave as waves. This also holds for light. In the nineteenth century, Maxwell showed that light consists of electromagnetic waves; and in 1905 Einstein described its particle nature as 'photons' — quanta of energy. The link with matter was made by de Broglie in 1924, when he showed by analogy that a particle has a wavelength associated with it by virtue of its motion.

The complementarity of these two descriptions is interwoven with the concept of coherence. Two or more waves can produce interference patterns when their amplitudes add up coherently. Interference phenomena, observed for both light and matter, embody many abstract concepts, including superposition of waves and the uncertainty principle.

For matter, superpositions analogous to those of light waves can be formed from matter wavefunctions. The Schrödinger equation yields wavefunctions together with their probability distributions, which are diffuse over position space. But if these waves are added up coherently with well-defined phases, the probability distribution becomes localized in space. The resultant wave packet and its associated de Broglie wavelength has the essential character of a classical particle: a trajectory in space and time with a well-defined (group) velocity and position — a

moving classical marble but at atomic scale!

The association of a wave character to particle motion through de Broglie's relationship is consistent with Heisenberg's mechanics, which define for all quantum systems an imprecision in position and in momentum. Similarly, there is an uncertainty relationship between time and energy. The joint effect of these two relations is the crux of the matter — short time is the way to localization. With femtosecond resolution, localization is on the atomic scale.

With such localization, a classical picture emerges. The atomic motions in any chemical or biological transformation are dynamic, with a speed of about 1 kilometre per second. So about 100 femtoseconds are needed to record atomic-scale dynamics over a distance of an angstrom. Freezing these motions as reactions unfold and pass through their transition states — the configurations between reactants and products — requires resolution in space and time, and only on this timescale (femtochemistry) can we directly observe the processes of breaking and making bonds.

If this is the case, why was it thought that femtosecond timescales would not be of value? There are two reasons. First, the energy uncertainty at this timescale was considered 'huge', and it was feared that the quantized vibrational-rotational states of molecules would be 'ruined' by using femtosecond resolution. Second, wave-packet spreading in microscopic systems was thought to be severe because of intra- and intermolecular interactions. Accordingly, femtosecond resolution would be of limited use for studies in chemistry and biology. Had the energy states been prepared incoherently, this would have been true, but this is not so. What is relevant is the energy uncertainty relative to the binding energy — for a 50-femtosecond duration, this ratio is about 1:70 for typical bonds.

Experimentally, atomic-scale dynamics of molecules and their reactions have now been observed and studied in all phases of

The uncertainty paradox

Femtosecond resolution is essential for observing atomic-scale dynamics in chemistry and biology. But quantum uncertainty was thought to be an obstacle — why wasn't it?

matter. One unique example is the breaking and remaking of bonds between sodium and iodine atoms. The wave packet oscillates periodically, showing particle-type behaviour during the whole reaction. The ensemble of molecules behaves in harmony, showing 'single-molecule' motion. The initial packet is indeed highly localized at about 0.1 Å and remains intact for many periods of its motion. The de Broglie wavelength is also consistent with sub-angstrom localization. This motion conforms to the coherent superposition and probing described above. Experiments in the condensed phase and in biological systems can be described similarly.

Why is coherence robust? To see motion in real systems, localized wave packets must form in every molecule, and there must also be a limited spread in position among the wave packets formed in the millions of molecules studied. This is achieved by the well-defined initial equilibrium configuration of the molecules before excitation and by the 'instantaneous' femtosecond launching of the packet. The spatial confinement of the initial ground state, typically 0.05 Å, ensures that all molecules, each with its own coherence, begin their motion in a bond-distance range much smaller than that of the actual motion, typically 5–10 Å. The femtosecond launching ensures that this narrow range of bond distance is maintained throughout preparation. Unless molecular and ensemble coherences are destroyed, the motion of the ensemble is that of a single-molecule trajectory.

Eugene Wigner and Edward Teller debated the uncertainty paradox for picosecond time-resolution in a lively exchange in 1972. But, because of coherence, the uncertainty paradox is not a paradox even for femtosience, and certainly not for the dynamics of physical, chemical and biological changes. Charles Townes encountered objections in the realization of the maser because of concern about the uncertainty principle, but coherence of photons was the key to success. As we cross the femtosecond barrier into the attosecond regime for studies of electron dynamics, we must recall this vital role of coherence; otherwise the spectre of quantum uncertainty might veil the path to new discoveries. ■

Ahmed H. Zewail is in the Chemistry and Physics Divisions, California Institute of Technology, Pasadena, California 91125, USA.



CORBIS

New-age drug meets resistance

Frank McCormick

A new sort of anticancer drug, designed to suppress the wayward protein that causes a type of leukaemia, is under the spotlight. Biological detective work has now revealed how this cancer becomes resistant to the drug.

Gleevec is the anticancer drug that everybody has been talking about. Known during its development as STI-571, this drug was rushed to market so quickly that Novartis, the company that developed it, had no time to clear a new trade name. 'Gleevec' was available — the name had been cleared for a different type of drug that didn't make it to market — so this trade name was slapped on STI-571. In the United States, Gleevec was approved for the treatment of chronic myeloid leukaemia (CML) in record time because of its sensational success in clinical trials. In one study, 53 out of 54 patients suffering from the early, 'chronic' phase of the disease showed complete, lasting responses¹. Patients suffering from the later stage, referred to as blast crisis, also responded well at first². But most relapsed within a few months, despite continued treatment.

Now the cause of the relapse has been identified. Writing in *Science*, Gorre and colleagues³ report that blast-crisis cells become resistant to STI-571, as a result of either a single mutation in the target protein — an enzyme known as p210 BCR-ABL — or the multiplication of the *BCR-ABL* gene. The results will have a direct effect on the way in which Gleevec is used to treat patients with CML. But they also have broader significance. Gleevec is at the forefront of a new wave of cancer treatments that are being designed rationally to block proteins directly involved in the diseases. The hope is that such drugs will act selectively against cancer cells and have few toxic side-effects.

The enzyme p210 BCR-ABL has long been considered an ideal drug target. It is a unique biochemical entity, encoded by the Philadelphia chromosomal translocation, which fuses part of the *BCR* gene from human chromosome 22 to part of the *c-ABL* gene on chromosome 9. The result is a tyrosine kinase — an enzyme that adds phosphate groups to tyrosine amino acids in target proteins — that is turned on all the time, and has biochemical properties distinct from those of the proteins from which it is derived⁴. In effect, the mutant enzyme drives the cells that contain it to become cancerous (Fig. 1); STI-571 works by turning it off.

The perfect anti-CML drug would hit p210 without affecting *c-ABL* (a tyrosine kinase in its own right) or other cellular targets. There are few other cancers with such a specific drug target⁵. However, STI-571 is

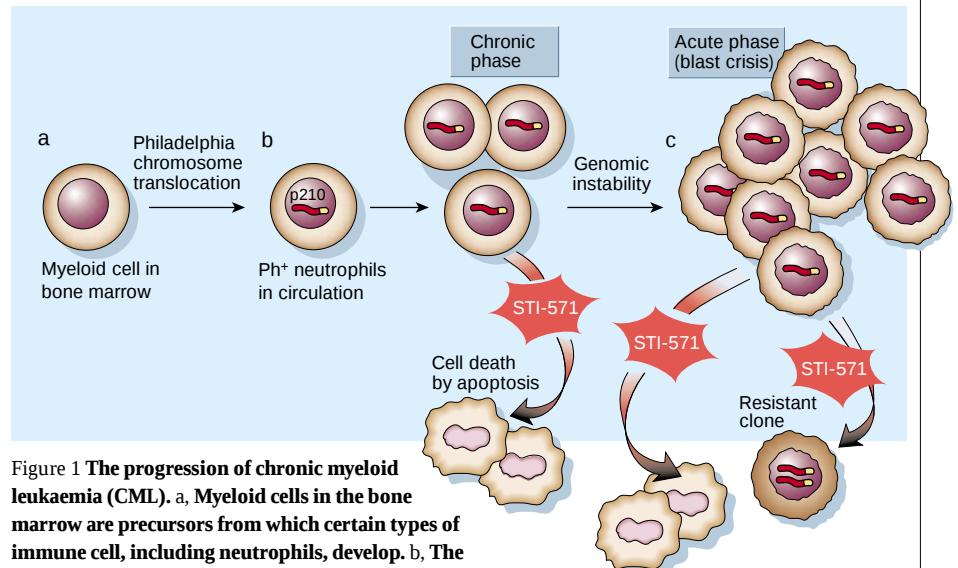


Figure 1 The progression of chronic myeloid leukaemia (CML). a, Myeloid cells in the bone marrow are precursors from which certain types of immune cell, including neutrophils, develop. b, The Philadelphia chromosomal translocation (Ph⁺) marks the onset of the early, chronic phase of CML. Here, part of the *BCR* gene becomes fused to the *ABL* gene. The result is the p210 BCR-ABL protein, which is turned on all the time and drives the cells towards cancer. The drug STI-571 (also known as Gleevec) has proved very successful in treating this phase of CML¹; it binds to and switches off p210. c, Further genomic instability (often including the loss of the tumour-suppressor proteins p53 and p16) pushes the cells towards the acute phase, 'blast crisis'. Patients at this stage initially responded well to STI-571, but most relapsed². Relapse results from the blast-crisis cells becoming resistant to STI-571, through a single mutation in p210 or an increase in *BCR-ABL* gene number³.

not a perfect drug in this respect: it does not discriminate between p210 and *c-ABL*, and it also inhibits other tyrosine kinases, such as the cell-surface receptor (*c-KIT*) that detects a protein known as stem-cell factor. This raised early fears of adverse side effects. Fortunately, these fears were unwarranted, and now Gleevec is also being tested for the treatment of a form of gastrointestinal cancer that is driven by *c-KIT*. These observations have effectively narrowed the criteria in the search for drugs specific to unique targets, and emphasize the fact that signal-transduction pathways are, to a certain extent, clearly redundant if several kinases can be inhibited in normal cells without having an effect.

The drawback comes with the discovery by Gorre *et al.*³ that patients with blast-crisis CML relapsed because their cancers became resistant to STI-571. The authors found that p210 was reactivated in blast-crisis cells from the patients who had relapsed. Analysis of the DNA from these cells showed that many patients had a single mutation in the *BCR-ABL* gene that rendered the protein product able to function in the presence of STI-571.

Others had many copies of the *BCR-ABL* gene, effectively meaning that their cells had more p210 than the drug could cope with.

Gleevec differs from other existing chemotherapies because it affects a protein that directly causes cancer — drugs such as 5-fluorouracil (5-FU) and methotrexate block enzymes that are more indirectly involved (thymidylate synthase and dihydrofolate reductase, respectively). Yet, despite this fundamental difference, resistance to Gleevec is depressingly similar to resistance to 5-FU and methotrexate. Single mutations in the target enzyme or increases in gene-copy number are often seen, as are other mutations that are less direct. For example, resistance to 5-FU can occur through alterations in other enzymes that affect the metabolism of fluorouracil⁶. In principle, resistance to STI-571 could also have arisen indirectly, and studies of more patients might well uncover more mechanisms. But so far it seems that resistance is acquired in such a way that p210 BCR-ABL is reactivated at full strength, underscoring the continued need for this protein in the progression of CML.

This is also not the first time that resistance to a drug with a unique protein target has been seen. Experience with acute promyelocytic leukaemia (APL) showed us it was possible. Like CML, APL results from chromosomal translocations (reviewed in ref. 7). This cancer is treated with all-*trans*-retinoic acid, which targets the protein that causes the disease — the PML-RAR α fusion protein. In this case, the treatment was identified before the molecular target⁸; nevertheless, this was one of the first examples of a cancer treatment with a specific target. Almost all APL patients respond to treatment with all-*trans*-retinoic acid, but resistance usually develops, most often through the mutation of RAR α .

Although not surprising, however, the findings by Gorre *et al.*³ raise several questions. Most obviously, why do drug-resistant cells emerge from blast crisis, and not from the earlier phase? Perhaps with time and more clinical data, this distinction will become blurred. However, cells from these phases are quite different. The chronic phase consists of a type of white blood cell (neutrophil) that is no longer dividing and is completely specialized. By contrast, the multiplication and dissemination of unspecialized myeloid or lymphoid precursor cells (from which many different types of immune cell usually develop) characterize blast crisis. These cell populations seem to respond with similar sensitivity to STI-571, at least in culture⁹. But mutations might occur more frequently in blast crisis simply because there are more dividing cells at this stage, or because of differences in DNA repair and genome stability.

Indeed, blast crisis is characterized by changes in the number of many different genes and by chromosome abnormalities, as well as by the loss of familiar tumour-suppressor proteins such as p53 and p16. Loss of p53 enables gene multiplication, and it will be interesting to see whether the amplification of *BCR-ABL* occurs more frequently in leukaemias lacking p53. However, there is no obvious molecular mechanism to explain the appearance of single mutations in p210 in blast-crisis cells.

For the future, we must hope to develop anticancer drugs with more sustained effects. By analogy with treatments for AIDS, cocktails of drugs — each with a different mechanism of action — might be more effective at achieving complete remission than would serial treatment with one drug after another, because the latter protocol would give the tumour the best chance of developing resistance to each drug. Several drugs that hit different parts of the same target might be ideal.

The experience with Gleevec also reinforces another message well known to oncologists: the earlier that cancer can be treated, the better. The effects of Gleevec on the chronic phase of CML are astounding and

are likely to reduce the frequency with which patients develop blast crisis. Unfortunately, even the earliest stages of colon cancer and breast cancer that can be detected with current technology have already undergone many genomic changes¹⁰, from this perspective, they resemble the blast-crisis phase of CML more than the chronic phase. New methods for detecting early stages of cancer will be vital, as will clinical trials that allow cocktails of drugs to be tested, in the hope of reducing the rate at which the inevitable enemy — drug resistance — develops. ■

Frank McCormick is at the UCSF Cancer Center, 2340 Sutter Street, San Francisco, California 94115, USA.

e-mail: mccormick@cc.ucsf.edu

1. Druker, B. J. *et al.* *New Engl. J. Med.* **344**, 1031–1037 (2001).
2. Druker, B. J. *et al.* *New Engl. J. Med.* **344**, 1038–1042 (2001).
3. Gorre, M. E. *et al.* *Science* **291**, 2091–2096 (2001).
4. Konopka, J. B., Watanabe, S. M. & Witte, O. N. *Cell* **37**, 1035–1042 (1984).
5. McCormick, F. *Curr. Opin. Biotechnol.* **11**, 593–597 (2000).
6. Allegra, C. & Grem, J. L. in *Cancer: Principles and Practice of Oncology* (eds DeVita, V. T., Hellman, S. & Rosenberg, S. A.) 432–448 (Lippincott-Raven, Philadelphia, 1997).
7. Melnick, A. & Licht, J. D. *Blood* **93**, 3167–3215 (1999).
8. Huang, M. E. *et al.* *Blood* **72**, 567–572 (1988).
9. Gambacorti-Passerini, C. *et al.* *Br. J. Haematol.* **112**, 972–974 (2001).
10. Gray, J. W. & Collins, C. *Carcinogenesis* **21**, 443–452 (2000).

Web link

www.gleevec.com

Astronomy

Dark arc casts a long shadow

John M. Dickey

The interstellar medium is mainly hydrogen gas, which is thought to cool to form clouds of molecular hydrogen. But the discovery of a large cloud of cold atomic hydrogen overturns that belief.

Interstellar clouds are immense collections of dust and gas floating between the stars in the Milky Way. These clouds come in different types, which are classified according to well-established theories of the evolution of the interstellar medium. But now, as reported on page 308 of this issue¹, Knee and Brunt have discovered a new kind of cloud in data from the Canadian Galactic Plane Survey. The structure they found is made of atomic hydrogen so cold (about 10 K) that it is visible only because it absorbs background radiation — essentially throwing a dark shadow across the sky. This ‘dark arc’ is impressive not only for its size, age and mass, but because according to conventional wisdom it should not exist at all.

Hydrogen is the most abundant element in the interstellar medium, and it is usually detected by its emission or absorption of radiation at specific wavelengths (spectral lines). It is now 50 years since Ewen and Purcell² detected radiation at a wavelength of 21 cm from cosmic atomic hydrogen — the first discovery of a spectral line at radio frequencies in space. This line, plus the spectral lines from molecules such as CO and OH at radio and millimetre wavelengths that were found afterwards, have had as profound an impact on our understanding of the interstellar medium as the double helix, discovered two years later, has had on genetics. Hydrogen is found everywhere in space, so mapping the radio emission at 21 cm allows us to trace the really big structures in the Milky Way.

The Canadian Galactic Plane Survey^{3–5} (CGPS) is one of several projects that are mapping the amount of atomic hydrogen in

our Galaxy by using the 21-cm emission line. These surveys are ambitious in combining data from interferometer arrays and single-dish radiotelescopes to achieve resolution that is an order of magnitude finer than any previous survey. The CGPS has already discovered several new structures in the outer Milky Way, including an immense chimney, through which hot gas from supernovae and stellar winds escapes from the main galactic disk⁶, and a giant mushroom cloud rising out of the disk⁷. The gas originating in the supernova is very hot and diffuse. As the remnant expands and cools over time, it forms a large shell. When shells overlap, or they run into small clouds, the gas density rises and new, larger clouds of atomic hydrogen are formed. The chimneys rain gas enriched in heavy elements down onto the disk of the Galaxy, where it rejoins the interstellar medium by trickling onto new clouds. The conventional picture is that atomic hydrogen cools and becomes denser, to the point where molecules can survive. After further cooling, new stars can form (Fig. 1).

There are also several small, very cold, atomic clouds that seem to contain no molecular hydrogen⁸. The newly discovered dark arc seems to be similar to these cold clouds, but much larger and more massive. Knee and Brunt¹ estimate that it is 6,000 light years across, with a mass at least 20 million times that of the Sun.

The trouble with these cold atomic clouds is that they do not fit into the dominant model^{9,10} for the interstellar medium, which was worked out in the late 1960s and early 1970s. As outlined above, this model has four phases, all of which are gaseous but

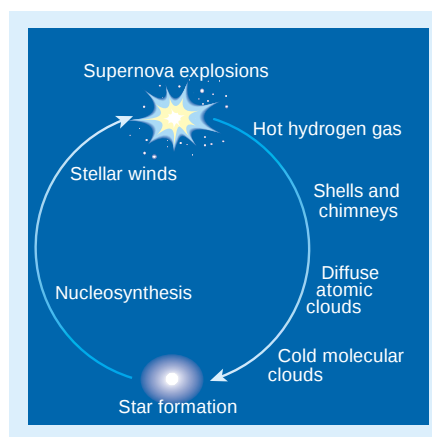


Figure 1 The galactic evolutionary cycle. Right, stars form from the raw material of hot hydrogen gas in the interstellar medium, which cools and becomes denser in passing around the right-hand side of the figure. The newly discovered dark arcs¹ belong on the lower right side, somewhere between the diffuse atomic and cold molecular clouds of hydrogen. But how they fit into the traditional understanding of the cycle is not clear. Left, star formation and evolution, ending in violent stellar death in supernovae explosions, and a return to the beginning of the cycle.

as different in density and temperature as ice, water and steam (Fig. 1). These traditional phases of the interstellar medium are the hot hydrogen gas (10^6 K) that fills old supernova remnants and stellar-wind bubbles; warm 'intercloud' gas (5,000–10,000 K) that contains both ionized and neutral hydrogen atoms; cool (50–100 K) 'diffuse' atomic clouds; and finally cold molecular clouds (10–25 K) in which almost all of the hydrogen is in the form of H_2 . The dark arc is as cold as a molecular cloud, but like the diffuse clouds it is made of atomic hydrogen.

Further, the arc does not look like a giant molecular cloud (GMC). For example, the molecular cloud in the Orion constellation includes bright concentrated regions (emitting at infrared wavelengths) that are probably close to gravitational collapse, young stars that have recently formed from the collapse of such regions, and bright spheres of ionized gas in regions of molecular hydrogen around these young stars. The dark arc shows none of these familiar traces of star formation, yet it must be tens of millions of years old — perhaps as old as 50 million years. This is very old for any kind of structure in the interstellar medium; the shearing motion caused by the rotation of the Galaxy would stretch the arc into oblivion if it were much older.

So what could have created this dark arc, and how can it stay so cold without becoming molecular? Did a large-scale process, such as the collision of gas with a spiral arm of the Milky Way, lead to its formation? Given its considerable age, it could even be the late stages of an expanding bubble of hot

gas formed by winds and supernova remnants from a cluster of young stars. In such cold, dense conditions, rates of molecule formation generally outpace the rate of photodissociation, which converts molecules back into atoms, so most gas in such clouds is molecular with only traces of atomic hydrogen. But as long as their density remains low there is nothing to prevent the formation of cold clouds of atomic hydrogen. Perhaps there are many more interstellar clouds like the dark arc out there, waiting to be discovered.

We do not yet know the origin of the dark arc. But in the meantime, others are finding similar clouds that are almost as cold. An immense atomic cloud discovered by astronomers¹¹ using the Green Bank Telescope in the United States is somewhat smaller and warmer than the dark arc, but is similar in having very little star formation. Perhaps these are examples of a new class of interstellar cloud, giant atomic clouds, which could rival the familiar GMCs in mass and size. It is telling that both of these clouds were observed only because of their self-absorption of radiation at 21 cm — a subtle feature in the emission spectrum that might not

have been noticeable except for the high angular and velocity resolution used, and careful removal of stray radiation in both of these studies. So perhaps it is not too surprising that these structures have been overlooked in the past. If the rate of these atomic cloud discoveries continues apace, astronomers will have to reconsider some of their ideas about the structure and evolution of galaxies.

John M. Dickey is in the Department of Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA.

e-mail: john@astro.umn.edu

1. Knee, L. B. G. & Brunt, C. M. *Nature* **412**, 308–310 (2001).
2. Ewen, H. I. & Purcell, E. M. *Nature* **168**, 356 (1951).
3. Taylor, A. R. in *New Perspectives on the Interstellar Medium* (eds Taylor, A. R., Landecker, T. L. & Joncas, G.) 3 (Astron. Soc. Pacific, San Francisco, 1999).
4. English, J. et al. *Publ. Astron. Soc. Australia* **15**, 56–59 (1998).
5. Normandeau, M., Taylor, A. R. & Dewdney, P. E. *Astrophys. J. Suppl.* **108**, 279–299 (1997).
6. Normandeau, M., Taylor, A. R. & Dewdney, P. E. *Nature* **380**, 687–689 (1996).
7. English, J. et al. *Astrophys. J. Lett.* **533**, L25–L28 (2000).
8. Gibson, S. J., Taylor, A. R., Higgs, L. A. & Dewdney, P. E. *Astrophys. J.* **540**, 851–862 (2000).
9. Field, G. B., Goldsmith, D. W. & Habing, H. J. *Astrophys. J. Lett.* **155**, L149–L154 (1969).
10. McKee, C. F. & Ostriker, J. P. *Astrophys. J.* **218**, 148–169 (1977).
11. Minter, A. H. & Lockman, F. J. *Am. Astron. Soc. Mtg* 198, Abstr. 59.07 (2001).

Signal transduction

An alternative to destruction

Daniel Finley

The traditional view is that, when a cellular protein becomes linked to a modifying group known as ubiquitin, it's curtains for that protein. But non-destructive functions for ubiquitin are now emerging.

The growth and survival of organisms depend on the transfer of information from the outside to the inside of cells. Outside the cell, molecules such as peptide hormones bind to specific receptors that sit in the cell membrane. This activates cascades of intracellular events, the earliest of which often involve chemical modification of regulatory proteins. Phosphate is a common type of modifying group. Another, about a hundred times larger, is ubiquitin¹. This protein is known to have a role in directing modified target proteins to a complex, the proteasome, that degrades them. This process might seem ill-suited to activate a signalling pathway, and indeed 'ubiquitination' often quenches biological signals by eliminating the proteins that convey the information. But Wang and colleagues' studies of the NF- κ B signalling pathway, described on page 346 of this issue², show that ubiquitination can also lead to the activation of an important enzyme, through a process that does not require protein degradation.

How can ubiquitin be used to activate a signalling process? Surprising answers to this question have already been provided by

studies of the NF- κ B family of transcriptional activators. In mammals, this protein family regulates many genes involved in inflammation, cell death and immunity.

NF- κ B has two subunits, one of which is generated from a precursor, p105. To activate transcription, this precursor must find its way to the nucleus, where transcription occurs; but p105 is kept outside the nucleus because its nuclear-localization sequence (an 'address label' that directs the protein to the nucleus) is masked by another part of the precursor. Ubiquitination of p105 leads to degradation of the masking element, exposing the nuclear-localization signal³. Here, then, ubiquitination of a protein can direct it to the proteasome to be activated rather than destroyed — provided that the protein can slip away from the proteasome after its self-inhibitory sequences have been removed. A variation on this processing mechanism, in which ubiquitin has a similar activating role, is provided by a membrane-anchored transcription factor that helps to regulate membrane fluidity⁴.

To reach the nucleus, NF- κ B must also escape its binding partner, I κ B. This happens

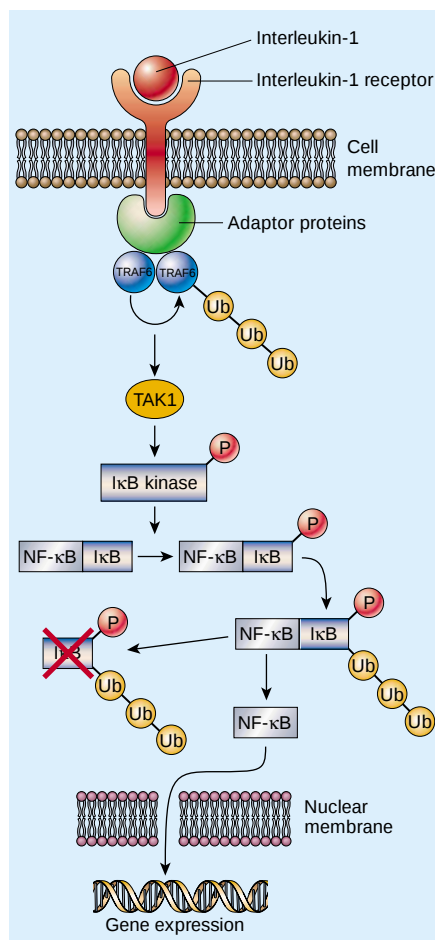


Figure 1 A non-destructive role for a traditionally destructive agent in signalling. Ubiquitin (Ub) is a small protein that, when linked to target proteins, usually signals their destruction. For example, in the NF- κ B signalling pathway, most of which is shown here, the destruction of the I κ B protein leads to the release of NF- κ B, which then moves to the nucleus where it can activate gene expression (lower part of the figure). Wang *et al.*² now reveal how ubiquitin can also have a non-destructive role in this signalling pathway. From the top, in response to the binding of the interleukin-1 molecule to its receptor, a set of 'adaptor' proteins (including MyD88 and IRAK) bind to the interleukin-1 receptor and recruit TRAF6 molecules, which are then modified by multiubiquitin chains. This somehow activates the TAK1 kinase, leading to the phosphorylation (represented by a circled 'P') of I κ B kinase, the phosphorylation and ubiquitination of I κ B, and subsequent events.

when I κ B is tagged with ubiquitin and then degraded, which occurs in response to signalling molecules such as tumour-necrosis factor- β and interleukin-1. So, in this case, ubiquitin mediates activation of the pathway by eliminating an inhibitor.

The trigger for the ubiquitination of I κ B is its phosphorylation by the I κ B kinase complex. And it was in the course of deciphering how this complex is itself switched on that Wang *et al.*² found a third example of how ubiquitin can activate signalling molecules. Here, ubiquitin has an early and non-destructive task.

Briefly, the results of Wang *et al.* point to the following model for one way of activating I κ B kinase (Fig. 1). First, interleukin-1 binds to the extracellular domain of the interleukin-1 receptor. Second, the intracellular part of the receptor recruits IRAK, a signalling molecule, which then binds to the TRAF6 protein. Third, this protein ubiquitinates itself. Fourth, ubiquitinated TRAF6 activates the kinase TAK1. And finally, this enzyme phosphorylates and activates I κ B kinase. The key finding is that a protein kinase — TAK1 — can be activated through ubiquitination of an associated protein. No protein degradation is needed: proteasomes were not present in the experiments², and nor do proteasome inhibitors affect the activation of I κ B kinase in cell extracts^{5,6} or *in vivo*³. So, ubiquitination

is used in three distinct ways to activate this pathway: it mediates processing of an NF- κ B precursor, destruction of an NF- κ B inhibitor, and non-destructive activation of TAK1.

Although it is unclear how ubiquitination of TRAF6 turns on the kinase activity of TAK1, this is the first non-destructive function for ubiquitin to be reconstituted biochemically using purified proteins. It seems that the self-ubiquitinating activity of TRAF6 is enough to drive the activation of TAK1: Wang *et al.*² show that the simple aggregation (oligomerization) of TRAF6 monomers, even in the absence of signalling from the interleukin-1 receptor, can trigger the ubiquitination of TRAF6 and subsequent events. The mechanism seems to resemble the activation of receptor tyrosine kinases, whereby receptor monomers that have oligomerized in the membrane phosphorylate each other.

Another protein that ubiquitinates itself is XIAP, which inhibits cell death⁷. Interestingly, this protein may also function through TAK1 to activate the phosphorylation of I κ B and the movement of NF- κ B into the nucleus⁸. This particular signalling pathway might begin with the activation of cell-surface receptors that bind bone morphogenetic proteins⁹, important during development. Ubiquitination of XIAP clearly targets it for degradation⁷, promoting cell death, but



100 YEARS AGO

A satisfactory trial of a new airship, devised by M. Santos Dumont, took place at Paris early on Saturday morning. It may be remembered that M. Henry Deutsch has offered a prize of 100,000 francs (4000l.) to the inventor of a flying machine which would travel from the heights of Saint Cloud to and round the Eiffel Tower and back again to the starting point in half an hour. M. Santos Dumont did the journey in forty-one minutes on Saturday, that is to say he made a voyage of nearly ten miles at a speed of about fifteen miles an hour, though the return journey was against the wind. He did not win the prize, but his attempt is very encouraging, for he has shown that an airship can be made to travel at a high speed in any direction. His airship consists of a cigar-shaped balloon, six metres in diameter, thirty-four metres long, and having a volume of 550 cubic metres.... M. Santos Dumont proposes to make another trial trip next Saturday. When going at full speed the propeller makes 200 revolutions per minute. From *Nature* 18 July 1901.

50 YEARS AGO

Our knowledge of the density (number and weight per sq. m.) of bottom animals in deeper waters is very scanty, since very few quantitative samples have hitherto been taken, and none at depths exceeding 1,000 – 1,400 m. It may therefore be of interest that the Danish Deep Sea Expedition in the *Galathea* has succeeded in taking thirteen quantitative samples with a 0.2 Petersen grab off the West African coast at depths between 820 m. and 3,782 m. ... the average weight of living animals in the six samples off the Congo from between 820 m. and 2,680 m. is 0.37 gm. per sample, that is, 1.85 gm. per sq. m. The number of species is on an average 2.5 per sample, that is 12.5 per sq. m., and the number of specimens 3 per sample, that is 15 per sq. m. These figures seem surprisingly high compared with those hitherto found at depths of about 1,000 m. in the Mediterranean, off Scotland and off Greenland. If the figures for the Angola coast are considered, the average weight per sample represents 0.5 gm., that is, 2.5 gm. per sq. m., with 3 species and 7 specimens per sample, that is 15 and 35 respectively per sq. m., which seems to be a very high figure considering the fact that the depths are between 1,000 m. and 4,000 m. From *Nature* 21 July 1951.

perhaps it has a transient signalling function before XIAP is destroyed.

Other examples of non-traditional functions for ubiquitin are sure to follow. There is already evidence that it has non-destructive roles in endocytosis (a process by which extracellular and cell-surface molecules are brought into cells)¹⁰ and the production of ribosomes¹¹ (biological machines that make proteins). And some functions of ubiquitin seem to be mediated through proteins that are not detectably degraded^{12,13}. As yet we do not understand the mechanisms underlying these events, but they present an interesting problem: how can ubiquitin signal the degradation of many target proteins and very different outcomes for others?

One clue might come from the fact that TRAF6 is ubiquitinated in an atypical way¹: several ubiquitin molecules are linked together into a chain through the lysine residue that each ubiquitin has at amino-acid position 63. In contrast, protein degradation is usually signalled by ubiquitin chains linked through lysine at position 48, and single copies of ubiquitin trigger the endocytosis of membrane proteins¹⁴. The linkage specificity for TRAF6 is set by two of the proteins that are involved in its ubiquitination⁵ — Ubc13 and Uev1A. The corresponding enzymes from yeast assemble to form a dimeric, but surprisingly asymmetric, complex¹⁵. Apparently, the last ubiquitin of the growing chain is docked at Uev1A in such a way that its lysine 63 residue is delivered directly to the active site of Ubc13. The next ubiquitin to be added to the chain can enter only from Ubc13, because only Ubc13 can chemically prime ubiquitin appropriately. So, the structural and functional asymmetries of the complex ensure that the chain has a unique topology¹⁵.

These insights show that enzymes of the ubiquitination pathway use highly specific

mechanisms to ensure that different chain structures or lengths are placed on target proteins, presumably to tailor different fates for the proteins. Yet there must be limits to the functional diversification that can be achieved through such mechanisms. The ubiquitin-like proteins¹⁶, with their variations in sequence, bypass these constraints. Protein modification by ubiquitin-like proteins has been implicated in numerous processes, including the transport of molecules from the cytoplasm to the nucleus, and from the cytoplasm to digestive organelles known as lysosomes. Surprisingly, none of these proteins has been found to direct its targets to the proteasome for degradation. Perhaps some of them have roles similar to the newly discovered function of ubiquitin, regulating the activity of their targets or of associated proteins. ■

Daniel Finley is in the Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA.

e-mail: Daniel_Finley@hms.harvard.edu

1. Weissman, A. M. *Nature Rev. Mol. Cell Biol.* **2**, 169–178 (2001).
2. Wang, C. et al. *Nature* **412**, 346–351 (2001).
3. Palombella, V. J., Rando, O., Goldberg, A. L. & Maniatis, T. *Cell* **78**, 773–785 (1994).
4. Hoppe, T. et al. *Cell* **102**, 577–586 (2000).
5. Deng, L. et al. *Cell* **103**, 351–361 (2000).
6. Chen, Z. et al. *Cell* **84**, 853–862 (1996).
7. Yang, Y., Fang, S., Jensen, J. P., Weissman, A. M. & Ashwell, J. D. *Science* **288**, 874–877 (2000).
8. Hofer-Warbinek, R. et al. *J. Biol. Chem.* **275**, 22064–22068 (2000).
9. Yamaguchi, K. et al. *EMBO J.* **18**, 179–187 (1999).
10. Hicke, L. & Riezman, H. *Cell* **84**, 277–287 (1996).
11. Finley, D., Bartel, B. & Varshavsky, A. *Nature* **338**, 394–401 (1989).
12. Robzyk, K., Recht, J. & Osley, M. A. *Science* **287**, 501–504 (2000).
13. Spence, J. et al. *Cell* **10**, 67–76 (2000).
14. Shih, S. C., Sloper-Mould, K. E. & Hicke, L. *EMBO J.* **19**, 187–198 (2000).
15. VanDemark, A. P., Hofmann, R. M., Tsui, C., Pickart, C. M. & Wolberger, C. *Cell* **105**, 711–720 (2001).
16. Jentsch, S. & Pyrowolakis, G. *Trends Cell Biol.* **10**, 335–342 (2000).

gen and its escape through the atmosphere and into space would result in a considerable overall loss of reducing power on the planet. The outcome would be that Earth's surface would eventually have become much more oxidized.

Mats of cyanobacteria (formerly called blue-green algae) consist of compact masses of microorganisms that grow on sediments and other solid surfaces in shallow aquatic environments (Fig. 1a)². The penetration of daylight into the mat is often limited to the top millimetre, yet the gross photosynthesis achieved, as measured from the rapid evolution of oxygen, is impressive — up to 5 grams of carbon per square metre per day. This is comparable to the production in tropical rain forests. The net productivity of mats, however, is marginal because 99% of the biomass produced is rapidly remineralized by bacteria (Fig. 1b). The mats are excellent food sources for snails, crustaceans and other small invertebrates, so nowadays they form persistent communities only in extreme environments such as salt ponds or hot springs, where the extreme conditions exclude grazing organisms. An extensive record of fossil microbial mats (stromatolites) is evidence that in Precambrian time — before about 600 million years ago and the evolution of multicellular grazers — similar communities apparently covered most environments that had the appropriate levels of water and sunlight³.

Today, hypersaline lagoons and intertidal flats such as those along the Pacific coast of Baja California, Mexico, provide ideal breeding grounds for cyanobacterial mats. Some, dominated by the species *Microcoleus chthonoplastes*, grow to many centimetres thick under the brine and have a finely laminated structure ranging in texture from jelly to rubber. Others, dominated by species of *Lyngbya*, persist on sand flats exposed to air, and form thin black carpets as tough as a doormat. Their coherent structure is created by intertwined microscopic filaments from billions of cyanobacteria embedded in gelatinous sheaths². To gain protection from the burning sun and from harmful ultraviolet radiation, the cells secrete sunscreen pigments, which make the mat appear almost black.

Hoehler et al.¹ studied hydrogen and carbon monoxide generation in both of these mat types, and discovered comparable physiological mechanisms of gas production in the two. The patterns of production were reversed, however, with high levels of carbon monoxide during the day, and high levels of hydrogen at night. The authors' larger argument hinges on hydrogen.

During the day, intense oxygen production in the millimetre-thin illuminated surface layer led to oxygen supersaturation and the development of permanent tiny gas bubbles trapped in the mats' fibres. The authors

Biogeochemistry

Space for hydrogen

Bo Barker Jørgensen

Mats of photosynthetic microorganisms known as cyanobacteria generate molecular hydrogen. If they were doing so earlier in Earth's history, the effect on the evolution of the atmosphere could have been profound.

Hydrogen is an ideal energy source, not only for an industrialized society but also for various common bacteria. In the anoxic world of soils and sediments, molecular hydrogen is generated by the bacterial degradation of organic material. The hydrogen is immediately recycled and supplies much of the reducing power required for the biological formation of vast amounts of methane and hydrogen sulphide.

On page 324 of this issue¹, Hoehler and colleagues show that mats of photosynthetic

cyanobacteria growing on coastal mud flats are hot-spots of biological hydrogen production. Although the process is indirectly driven by solar energy, it takes place mainly at night when oxygen levels are low. The authors draw the astonishing conclusion that such a process might have contributed to the oxidation of the primordial Earth at a time when ancestors of modern microbial mats dominated the biosphere. Their reasoning is that, over time spans of a billion years or so, even a slow generation of hydro-

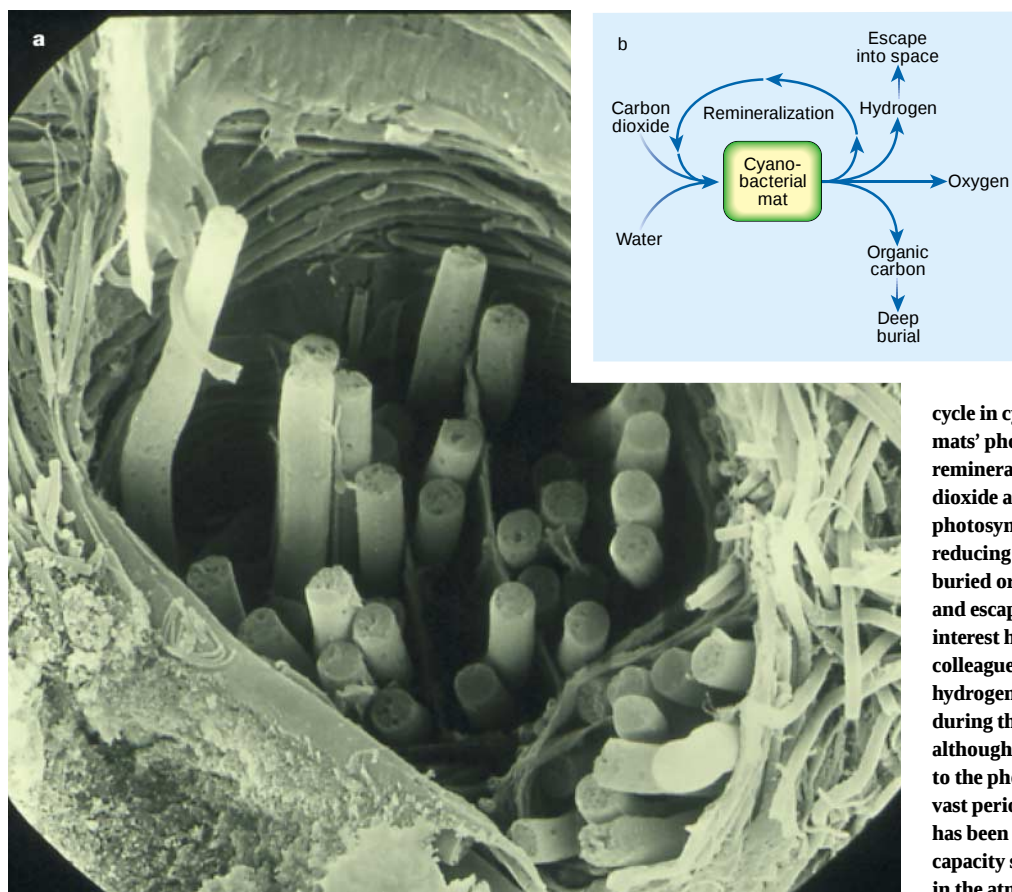


Figure 1 Cyanobacteria, and their possible influence on Earth's early atmosphere. **a**, Scanning electron micrograph of spaghetti-like, 3- μ m-wide filaments of the microorganism *Microcoleus chthonoplastes*. Cyanobacteria such as these form dense mats that grow in shallow aquatic environments. **Before the evolution of multicellular grazing organisms, such mats may have constituted the dominant biological community on Earth.** **b**, The biogeochemical

cycle in cyanobacterial mats. Almost all of the mats' photosynthetic biomass production is remineralized by bacteria back into carbon dioxide and water. The oxygen resulting from photosynthesis can accumulate only when reducing power in the form of organic carbon is buried or when molecular hydrogen is generated and escapes into space. The feature of especial interest here is the discovery by Hoehler and colleagues¹ that cyanobacterial mats produce hydrogen as a by-product of nitrogenase activity during the night. Their hypothesis is that although hydrogen production is small relative to the photosynthesis–respiration cycle, over vast periods of geological time enough hydrogen has been lost to space to lower Earth's reducing capacity significantly and so raise oxygen levels in the atmosphere.

applied a simple but elegant technique to sample such bubbles repeatedly and analyse their gas composition throughout the day and night. Surprisingly, hydrogen concentrations in the *Lyngbya* mats were ten thousand times higher late at night than during the day. In *Microcoleus* mats, the hydrogen levels were also highest at night. But the day–night variation was much less pronounced, perhaps because, unlike *Lyngbya*, *Microcoleus* does not produce gas-reservoir bubbles.

Hoehler *et al.* propose that, in addition to its formation from fermentation processes in the anoxic mat, hydrogen is generated by the cyanobacteria themselves as a side reaction of their nitrogenase enzyme system. Nitrogenase occurs in many microorganisms, its main function being to 'fix' molecular nitrogen and convert it into nitrogen-rich cellular proteins. Marine ecosystems generally suffer from nitrogen starvation, so this process provides a crucial extra source of the nutrient.

Nitrogenase might, however, also be used to channel electrons into protons to form planktonic hydrogen, a process known from planktonic cyanobacteria in lakes and in the sea⁴. The cyanobacteria are thereby relieved of excess reducing power accumulated during daylight in the form of storage carbohydrates in the cells. Nitrogenase is sensitive to oxygen, so most nitrogen-fixing cyanobacteria contain it in specialized cells,

heterocysts, that lack the oxygen-generating photosynthetic machinery and have thick walls impermeable to external oxygen. Non-heterocystous cyanobacteria lack this oxygen-protection mechanism and therefore require a spatial or temporal separation of photosynthesis and nitrogen fixation. As a result, their nitrogenase is active mainly during the night⁵. The results of Hoehler *et al.* show that biogeochemists likewise had to be active during the night to discover this transient phenomenon.

It is well known that, in anoxic environments, a large fraction of the entire energy flow from organic matter is channelled through hydrogen, which is taken up by microorganisms and used to generate methane or hydrogen sulphide. Why, then, is the hydrogen production of cyanobacterial mats so exciting? There are two reasons. First, some of the hydrogen generated at the surface of the mats may evade rapid recycling by bacteria and escape into the atmosphere. Second, these mats have been the dominant biological communities for most of Earth's history, and their catalytic capabilities might therefore have had a profound influence on the chemical development of our planet.

Although oxygen-generating photosynthesis might have evolved more than 2.7 billion years ago, the concentration of oxygen in the atmosphere did not exceed 1–2% of the modern level until half a billion years later^{6,7}. A net accumulation of oxygen on

Earth requires that an equivalent quantity of reduced chemical species be taken out of the global cycles and stored away for long geological periods. The accelerating build-up of an oxygenated atmosphere and ocean over the Proterozoic era, from about 2.2 billion years onwards, was thus balanced by a burial of organic carbon and pyrite (iron disulphide) deep in marine sediments⁶. Loss of biologically produced hydrogen from the outer atmosphere into space could have been another mechanism for the removal of reducing power. Until now, however, this route has been considered to be of minor importance: only hydrothermal vents and volcanoes, or the ultraviolet-catalysed oxidation of atmospheric methane to hydrogen and carbon monoxide, have been considered as hydrogen sources.

Hoehler *et al.*¹ propose that hydrogen generation by cyanobacterial nitrogenases might have been a further source, and one that had a global impact. Taking the results of their experiments on small pieces of mat from Baja California, they courageously extrapolate back to the first half of Earth's evolution. The gas budgets that they estimate are based on the assumptions that global primary productivity of microbial mats on the early Earth was comparable to that of modern marine ecosystems and that hydrogen emission rates were likewise similar to those studied. If the hydrogen were completely lost to space from an early atmosphere that was

nearly devoid of oxygen, then this sink for a reducing species could have been as important as the deep burial of organic carbon for oxygenating Earth's surface.

There are still huge uncertainties and many open questions, of course. Were the primordial mats physiologically similar to those of modern *Microcoleus* or *Lyngbya* communities? What extent of the Earth's surface did they cover, and how productive were they? When did nitrogenase first occur in the ancestors of modern cyanobacteria? Nonetheless, in principle Hoehler *et al.* show that over a vast period of time the release of even a biological trace gas from a metabolic side reaction might have had profound consequences. With this work we can add a new piece of

evidence to the great puzzle of early biogeochemical development on our planet. ■

Bo Barker Jørgensen is at the Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, D-28359 Bremen, Germany.

e-mail: bjoergen@mpi-bremen.de

1. Hoehler, T. M., Bebout, B. M. & Des Marais, D. J. *Nature* **412**, 324–327 (2001).
2. Cohen, Y. & Rosenberg, E. (eds) *Microbial Mats: Physiological Ecology of Benthic Microbial Communities* (Am. Soc. Microbiol., Washington DC, 1989).
3. Schopf, J. W. & Klein, C. (eds) *The Proterozoic Biosphere* (Cambridge Univ. Press, 1992).
4. Conrad, R. *Adv. Microb. Ecol.* **10**, 231–283 (1988).
5. Stal, L. J. *New Phytol.* **131**, 1–32 (1995).
6. Des Marais, D. J., Strauss, H., Summons, R. E. & Hayes, J. M. *Nature* **359**, 605–609 (1992).
7. Nisbet, E. G. & Fowler, C. M. R. *Proc. R. Soc. Lond. B* **266**, 2375–2382 (1999).

Immunology

End game for B cells

Kathryn Calame

Plasma cells are a key part of the immune system, secreting antibodies to ward off invaders. The discovery of a protein required for the production of plasma cells will give immunologists a better idea of how they develop.

The immune system has two lines of strategic defence against foreign invaders. The cellular response is provided by T cells, which kill virally infected or malignant cells. The humoral response consists of protective molecules (antibodies) that are released into the bloodstream. The cells that release antibodies — plasma cells — are specialized B cells that no longer divide and have reached the end of their programme of development. These cells are literally antibody factories, filled with a membranous protein-secretion apparatus (the endoplasmic reticulum) and dedicated to producing and secreting antibodies.

Because antibodies are involved in the progression of many autoimmune diseases — those in which an individual's own cells trigger an immune response — studies of autoimmunity would benefit greatly from a better understanding of plasma cells. Yet surprisingly little is known about these cells, even though the regulation of B-lymphocyte development in general has been heavily studied. On page 300 of this issue¹, Reimold *et al.* take us further with their discovery that a protein called XBP-1, which activates gene expression, is essential for plasma cells to develop and to secrete antibody.

XBP-1 activates genes by binding to a specific DNA sequence called an X box, first identified in a regulatory region of a gene needed for immune defence (a class II major histocompatibility gene)². XBP-1 is expressed in all tissues, and mice engineered to lack this protein die as embryos as a result of liver abnormalities and severe anaemia³. This makes it difficult to determine what the

protein does in specific cell types, such as lymphocytes, in adult animals. Yet, given their observations that the expression of XBP-1 is induced when splenic B cells are activated *in vitro* and is high in plasma cells *in vivo*¹, this is exactly what Reimold *et al.* wanted to do.

The authors¹ used a well-established system⁴ that takes advantage of the fact that RAG proteins are essential for early immune cells to rearrange their immunoglobulin (antibody) and T-cell-receptor genes, and to continue developing. The authors took embryonic stem cells — which are capable of developing into all cell lineages — from XBP-1-deficient mice, and injected them into RAG-deficient embryos at an early stage of development. Nearly all cell lineages in the resulting animals were chimaeric, consisting of some cells that contained RAG genes but lacked the *XBP-1* gene, and some that contained *XBP-1* but lacked RAG genes. However, because RAG activity is crucial for lymphocyte development, all T and B cells were derived from the embryonic stem cells that lacked *XBP-1* but could express RAG proteins normally.

The B and T cells in the bone marrow, thymus and spleen of the chimaeric mice appeared normal in their numbers and development. Furthermore, when XBP-1-deficient splenic B cells were stimulated *in vitro*, they showed normal activation, as judged by their expression of surface markers, proliferation, and rearrangement of the gene encoding one of the two polypeptides in an antibody molecule. But the cells failed to express Syndecan-1 (a surface marker for

antibody-secreting cells) and to secrete immunoglobulins in the usual way. Re-introducing XBP-1 *in vitro* restored the cells' ability to secrete antibodies, confirming that the defect resulted from lack of XBP-1.

The authors also looked at the ability of B cells in the chimaeric mice to secrete specific antibodies after exposure to polyoma virus or to various model antigens. Germinal centres — the specialized microenvironments in which B cells receive help from T cells, multiply and undergo further development — were normal. But all the chimaeric mice failed to secrete antibody into their bloodstreams, and Reimold *et al.* detected few plasma cells in the spleens. So, XBP-1 is not required for B cells to be activated or to develop through the germinal-centre stage. But it is needed for the final stage of B-cell development: the production of antibody-secreting plasma cells. XBP-1 is the first transcriptional regulator found to be required specifically for plasma-cell development.

This discovery¹ sheds light on some earlier observations of the regulation of XBP-1. Expression of the *XBP-1* gene is repressed by B-cell-lineage-specific activator protein (BSAP), which is essential for early B-cell development⁵. BSAP levels fall in plasma cells⁶, and the fact that XBP-1 is necessary for plasma-cell development offers one explanation for why this fall is important. Moreover, interleukin-6, another important regulator of plasma cells, induces the expression of XBP-1 in myeloma cells (cancerous plasma cells)⁷. Finally, the increased synthesis of antibody that occurs in plasma cells might put the endoplasmic reticulum under strain; endoplasmic-reticulum stress is known to activate a transcriptional regulator called ATF6, which in turn induces the expression of XBP-1 (ref. 8). If XBP-1 is needed for increased antibody production, then the possible increased expression of XBP-1 after endoplasmic-reticulum stress might provide a positive feedback loop. So, the fall in BSAP levels, interleukin-6-dependent signals and the activation of ATF6 may all be required to achieve maximal XBP-1 expression in plasma cells.

Other proteins — namely interferon regulatory factor-4 (IRF-4) and B-lymphocyte-induced maturation protein-1 (Blimp-1) — are also involved in plasma-cell development. IRF-4-deficient mice have defects in both T and B lymphocytes, and do not mount antibody responses⁹. Forced expression of Blimp-1 in B-cell lines and splenic B cells is sufficient to drive their development to an antibody-secreting state^{10,11}. Both factors are expressed in plasma cells and in a subset of B cells, found in the germinal centre, that have features of plasma cells^{12,13}.

How does XBP-1 interact with these proteins? It is not clear exactly when XBP-1 acts during plasma-cell differentiation, nor whether its levels are high in those germinal-

centre B cells that express IRF-4 and Blimp-1. However, Reimold *et al.*¹ found that activated XBP-1-deficient B cells, which do not develop into antibody-secreting plasma cells, do express Blimp-1. So this factor cannot substitute for XBP-1, which might act at a later stage of development than both Blimp-1 and, possibly, IRF-4. Reimold *et al.*'s results lay the groundwork for experiments to determine whether these three proteins act in the same or parallel regulatory pathways. They also provide a system for identifying the genes that are regulated by XBP-1 in plasma cells.

Reimold *et al.*'s unexpected finding¹ offers an excellent starting point for further studies of the development and function of plasma cells. Such information will shed light on the general mechanisms that are involved when cells commit to a terminally differentiated state, as well as those processes specific to how B lymphocytes develop into plasma cells. A better knowledge of plasma-cell biology should also help in understanding certain human diseases. For example, it might provide insight into how plasma cells

become deregulated and hence cancerous, leading to multiple myeloma, and why antibodies that recognize self proteins are produced inappropriately in autoimmune disorders such as myasthenia gravis, systemic lupus erythematosus and Graves' disease. ■

Kathryn Calame is in the Department of Microbiology, Columbia University College of Physicians and Surgeons, 701 West 168th Street, New York, New York 10032, USA.
e-mail: klc1@columbia.edu

1. Reimold, A. M. *et al.* *Nature* **412**, 300–307 (2001).
2. Liou, H. C. *et al.* *Science* **247**, 1581–1584 (1990).
3. Reimold, A. M. *et al.* *Genes Dev.* **14**, 152–157 (2000).
4. Chen, J., Lansford, R., Stewart, V., Young, F. & Alt, F. W. *Proc. Natl Acad. Sci. USA* **90**, 4528–4532 (1993).
5. Nutt, S. L., Heavey, B., Rolink, A. G. & Busslinger, M. *Nature* **401**, 556–562 (1999).
6. Barberis, A., Widenhorn, K., Vitelli, L. & Busslinger, M. *Genes Dev.* **4**, 849–859 (1990).
7. Wen, X. *et al.* *Int. J. Oncol.* **15**, 173–178 (1999).
8. Yoshida, H. *et al.* *Mol. Cell. Biol.* **20**, 6755–6767 (2000).
9. Mittrucker, H. W. *et al.* *Science* **275**, 540–543 (1997).
10. Piskurich, J. F. *et al.* *Nature Immunol.* **1**, 526–532 (2000).
11. Turner, C. A. Jr, Mack, D. H. & Davis, M. M. *Cell* **77**, 297–306 (1994).
12. Falini, B. *et al.* *Blood* **95**, 2084–2092 (2000).
13. Angelin-Duclos, C., Cattoretti, G., Lin, K. I. & Calame, K. *J. Immunol.* **165**, 5462–5471 (2000).

Superconductivity

Iron cast in exotic role

S. S. Saxena and Peter B. Littlewood

Conventional wisdom says that superconductivity and magnetism are incompatible bedfellows. So the idea of iron as a superconductor is ruled out — or is it?

Iron is one of the commonest heavy elements on Earth and is the best known example of a magnetic metal. Up to temperatures of about 1,000 K the stable structure of iron is that of a body-centred cubic (bcc) crystal, and it is strongly ferromagnetic — that is, capable of being permanently magnetized (Fig. 1). In its familiar magnetic state, the quantum-mechanical spins of the electrons around the iron atoms in the bcc lattice spontaneously align parallel to each other; this parallel alignment produces the bulk ferromagnetism. Most of the Earth's iron is in its core. Under the high-pressure conditions in the solid inner core, iron has a different stable structure — that of a hexagonal close-packed (hcp) crystal (Fig. 1). In this high-pressure phase the ferromagnetism of iron is believed to disappear, and it has been predicted^{1,2} that in this non-magnetic state iron could become a superconductor. On page 316 of this issue, Shimizu *et al.*³ report the first evidence for superconductivity in iron at very low temperatures and high pressures.

To those accustomed to thinking of iron as the archetypal strong ferromagnet, this discovery will seem particularly remarkable. Superconductivity occurs in conventional

non-magnetic metals when they are cooled to low enough temperatures for their conduction electrons to pair up and flow without resistance. However, tiny magnetic impurities can destroy superconductivity in conventional superconductors by breaking up the electron pairs. And even when true magnetic order does not occur, the dynamic magnetic fluctuations that occur when a material is on the verge of ferromagnetism can be enough to suppress superconductivity, as in palladium⁴. In addition to having zero electrical resistance, superconductors expel magnetic fields from their interior — the so-called Meissner effect. (This dramatic effect allows a permanent magnet to be levitated above a superconductor.) So it has long been thought that superconductivity and ferromagnetism are incompatible. But despite these widespread beliefs, superconductivity and ferromagnetism can mix when the magnetization is small^{5,6}.

In its ferromagnetic bcc state, even at very low temperatures, superconductivity has never been observed in iron. This is where the high pressures come in. Applying pressure to iron changes its structure to hcp, thereby destroying its ferromagnetic order. In this situation superconductivity becomes

possible. Shimizu *et al.*³ show that this occurs for iron at a pressure above 10 gigapascals (about 100,000 times atmospheric pressure) and at temperatures below 2 K. Remarkably, they even observe the Meissner effect in their experiment, confirming that iron reaches a true superconducting state at high pressures.

Although ferromagnetism is suppressed in the hcp phase, there is a growing body of evidence that the hcp phase of iron is weakly antiferromagnetic at low temperatures (ref. 7; L. Vocadlo and D. Alfè, personal communication). In an antiferromagnetic state, the electron spins on alternative atoms are aligned antiparallel to each other, rather than parallel, and sometimes in more complicated arrangements. In the simplest case, when half the spins point in one direction and the other half in the opposite direction, there is no net magnetization.

What could be the mechanism of superconductivity in iron? As with all new superconductors the origin of the electron pairing — the attraction that allows the electrons to overcome their usual repulsion — is the main issue to be resolved. In conventional superconductors, such as most elemental metals, elastic vibrations (phonons) in the crystal lattice mediate the pairing between electrons. In iron, a phonon mechanism is a strong possibility, but the type of superconductivity that is favoured by this mechanism is easily suppressed by strong magnetic fluctuations. So if this is the correct model, the magnetic fluctuations must be quite weak, a result that has implications for the Earth's

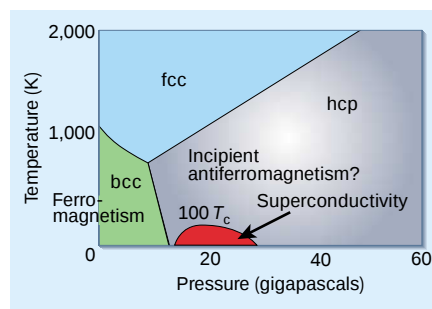


Figure 1 The temperature–pressure phase diagram of iron. The low-pressure and low-temperature phase of iron has a body-centred cubic (bcc) structure and is strongly ferromagnetic. The higher-pressure phase has a hexagonal close-packed (hcp) structure and might be weakly or nearly antiferromagnetic at low temperatures. Shimizu *et al.*³ show that below 2 K and at pressures above 10 GPa the hcp phase of iron becomes superconducting. (Here the superconducting transition temperature, T_c , has been magnified by a factor of about 100.) The shape of the phase diagram at low temperatures in the border region between the ferromagnetic bcc and the superconducting hcp phases has still not been established experimentally. The third structural form of iron shown here is fcc (face-centred cubic).

dynamo — the origin of our planet's magnetic field. It is fluid motion in the Earth's liquid outer core that generates the magnetic field. But just as in the core of an electromagnet, the properties of the solid inner core can damp out the fast fluctuations and stabilize the geodynamo — inhibiting field reversals, for example, which occur historically a few times every million years. Of course the solid core is much too hot to be superconducting. But the dynamo can be stabilized both by a large electrical conductivity and by strong paramagnetic fluctuations⁸ (paramagnetism is a weak sort of magnetism that exists only when there is an applied magnetic field). Without strong paramagnetic fluctuations, one mechanism⁸ for stabilizing the geodynamo against field reversals is absent.

If the magnetic fluctuations (either anti-ferromagnetic or paramagnetic) are in fact large, an alternative possibility is that the superconductivity discovered by Shimizu *et al.*³ might be mediated by magnetic spin fluctuations. A magnetic pairing mechanism is likely to produce a superconducting state with characteristics different from those of the phonon-paired state, which should be testable experimentally. If this view holds, superconductivity in high-pressure iron might be related to that found in a group

of unconventional superconductors that includes the heavy-fermion compounds and the copper-oxide high-temperature superconductors. Such an exotic mechanism for superconductivity in hcp iron, one requiring strong magnetic fluctuations, might fit in better with a geological model of a strongly paramagnetic core.

Deciding between the two types of mechanism — conventional phonon-mediated and exotic magnetically mediated — will require further measurements. But, unusually, the outcomes of these measurements will provide fresh information not just for theories of superconductivity but also about Earth's geomagnetism. ■

S. S. Saxena is in the Department of Physics, University College London, Gower Street, London WC1E 6BT, UK.

Peter B. Littlewood is in the Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK. e-mail: pbl21@phy.cam.ac.uk

1. Wohlfarth, E. P. *Phys. Lett.* **75A**, 141–143 (1979).
2. Freeman, A. J., Continenza, A., Massida, S. & Grossman, J. C. *Physica C* **166**, 317–322 (1990).
3. Shimizu, K. *et al. Nature* **412**, 316–318 (2001).
4. Doniach, S. *Proc. Phys. Soc. (Lond.)* **91**, 86 (1967).
5. Saxena, S. S. *et al. Nature* **406**, 587–592 (2000).
6. Pfeleiderer, C. *et al. Nature* **412**, 58–61 (2001).
7. Steinle-Neumann, G., Stixrude, L. & Cohen, R. E. *Phys. Rev. B* **60**, 791–799 (1999).
8. Gilder, S. & Glen, J. *Science* **279**, 72–74 (1998).

Cell biology

Cytoskeleton in the cell cycle

Yukinobu Nakaseko and Mitsuhiro Yanagida

During cell division, a series of checkpoints make sure that events are proceeding normally. The latest checkpoint to be discovered monitors the actin cytoskeleton, ensuring that it is organized correctly.

The internal filamentous structures that support a cell and give it shape are more dynamic than their name — the cytoskeleton — might suggest. The cytoskeleton is constantly changing and rearranging, contributing to a variety of cellular functions. The two main components of the cytoskeleton are actin, which forms polymers known as thin filaments, and tubulin, which polymerizes to form microtubules. Microtubules are crucial to cell division: they form the spindle, which segregates duplicated chromosomes into two new cells. But, even though the actin-based cytoskeleton also undergoes dramatic changes throughout the cell-division cycle^{1,2}, it has been unclear whether actin is involved in cell division. Writing on page 352 of this issue³, Gachet and colleagues tackle the problem. It seems that cells keep a check on the actin cytoskeleton, and, if it is defective, the spindle becomes orientated incorrectly, and subsequent phases of cell division are suppressed.

An early phase of the cell-division cycle involves the duplication of the cell's genetic

material. Later, during mitosis (nuclear division), spindle microtubules attach to the 'kinetochore' regions on the duplicated chromosomes (the sister chromatids). The spindles elongate, pulling the chromatids apart to opposite poles of the nucleus (Fig. 1a, overleaf), which then divides into two. At the molecular level, this process is rather complicated. For example, a protein called Rad21 keeps the sister chromatids glued together during metaphase — the stage that marks the middle of mitosis⁴. At the start of the next stage, anaphase, the anaphase-promoting complex activates the degradation of the securin protein. Once degraded, securin can no longer inhibit a third protein, separase, leaving it free to degrade Rad21 (ref. 4). This means that the sister chromatids can now separate.

If the spindle fails to attach to the chromosomes, cells stop or delay cell division during metaphase, so the proper connection can be re-established (Fig. 1c). This insurance system, called the spindle-assembly checkpoint, is evolutionarily conserved, and

involves many different proteins⁵. For instance, the Mad2 protein senses unattached kinetochores, and restrains mitosis by inhibiting the anaphase-promoting complex⁶. This inhibits the degradation of securin and subsequent events, as well as the degradation of another protein, cyclin B, which is involved in mitotic control.

The novelty in Gachet *et al.*'s work³ is their finding that the actin cytoskeleton is also important in mitosis, being implicated in a checkpoint — unheard of until now — that ensures the correct segregation of chromosomes (Fig. 1b). Actin's involvement in cell motility is well known, but its role in cell-cycle control has been less well understood. Gachet *et al.*'s initial discovery³ was that a drug that inhibits actin polymerization, latrunculin B⁷, causes fission yeast to delay nuclear division. Cells treated with latrunculin B stop during mitosis for about an hour; the spindle microtubules remain short, sister chromatids are not separated, and nuclei do not divide. Moreover, the axis of the spindle is often orientated incorrectly. The authors observed similar features in cells with a mutation in actin, and in cells treated with another inhibitor of actin polymerization. So, disorganization of the actin cytoskeleton seems to perturb proper spindle orientation, which apparently activates a new type of checkpoint system that delays mitosis.

More specifically, the authors show that this checkpoint blocks the normal separation of sister chromatids. They find that, in cells that are treated with latrunculin B, the breakdown of Rad21 is delayed compared with untreated cells. Moreover, if Rad21 is mutated, the cells do not arrest during mitosis in the presence of latrunculin B — they die instead.

As mentioned above, the spindle-assembly checkpoint also signals to Rad21, among other proteins. Surprisingly, though, the authors find that the new actin-dependent checkpoint is otherwise independent of the spindle-assembly checkpoint (Fig. 1). In cells with disrupted Mad2, the actin-dependent checkpoint is normal. Moreover, in wild-type cells treated with latrunculin B, the oscillation of activity of the mitotic protein Cdc2 kinase is also normal, as is the timing of destruction of cyclin B and securin. In other words, the actin-dependent checkpoint delays the separation of sister chromatids without activating the spindle-assembly checkpoint, and without inhibiting the anaphase-promoting complex.

What, then, are the cellular factors that activate this new mitotic checkpoint? Gachet *et al.* addressed this question by searching for mutant cells that could not grow in the presence of a low concentration of latrunculin B. They found that cells lacking the proteins Atf1 or Sty1/Spc1 were hypersensitive to the drug. Sty1/Spc1 is a mitogen-activated pro-

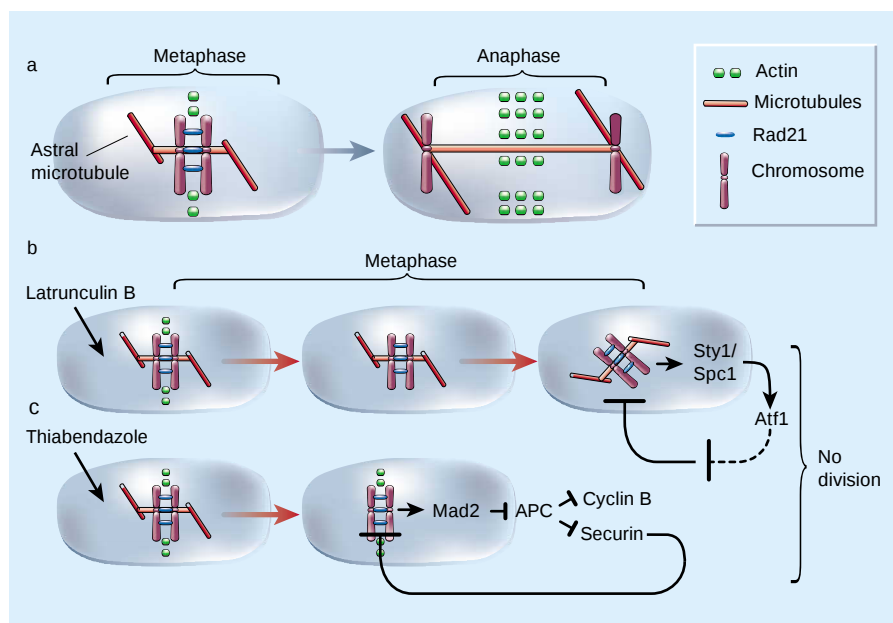


Figure 1 Keeping a check on the cytoskeleton. a, During nuclear division, spindle microtubules attach to duplicated chromosomes (in metaphase) and pull them apart to opposite poles of the nucleus (in anaphase). b, As shown by Gachet *et al.*³, when the actin-based cytoskeleton is damaged by treatment with a drug, latrunculin B, progression into anaphase is halted. The molecular details of this new cell-cycle checkpoint have not been fully worked out, but actin disorganization somehow results in the spindle being misorientated, which activates the proteins shown. The ultimate result is a failure to degrade Rad21 and to segregate duplicated chromosomes. c, The 'spindle-assembly' checkpoint can be triggered by thiabendazole, a drug that damages tubulin. When chromosomes detect that they are no longer attached to microtubules, the spindle-assembly checkpoint is activated. This involves different molecules from those in the actin checkpoint, except that one of the ultimate targets is also Rad21.

tein kinase, which adds phosphate groups to Atf1, activating it. Atf1 is a transcription factor, working to control gene expression. Both Atf1 (ref. 8) and Sty1/Spc1 have counterparts in humans. The authors found that cells lacking Atf1 failed to arrest during mitosis in the presence of latrunculin B, and later died. The loss of actin polymerization (but not the addition of a tubulin poison) induced the phosphorylation and activation of Sty1/Spc1. Gachet *et al.* conclude that the 'stress-activated' protein kinase pathway, which includes Sty/Spc1, is a key part of the actin-dependent mitotic checkpoint.

To explain how actin affects spindle orientation, Gachet *et al.* propose that there is a functional connection between astral microtubules, which extend outwards from the spindle, and the ring of actin that forms in the cell cortex (the outermost layer of the cytoplasm) during mitosis. The actin ring might use the astral microtubules to orientate the spindle microtubules. If so, when the actin ring becomes disorganized, the spindle would be orientated incorrectly. The stress-activated pathway would then be switched on, so the sister chromatids would fail to separate. This is still hypothetical, particularly because the role of astral microtubules in organizing spindle microtubules is unclear. Nevertheless, it seems plausible: the interaction of microtubules with the cortex is

known to be important in positioning dividing nuclei⁹.

This work might have broad implications. For example, cells acquiring mutations that enable them to bypass this 'spindle-orientation' checkpoint might develop genomic instability — a feature of tumour cells. Moreover, spindle orientation is known to be important in determining the fate of many cell types during the early development of multicellular animals¹⁰. Specifically, the orientation of the spindle is involved in whether a cell divides symmetrically or asymmetrically. So the new checkpoint might also be linked to the control of cell fate.

Yukinobu Nakaseko and Mitsuhiro Yanagida are at the CREST Research Project, Department of Gene Mechanisms, Graduate School of Biostudies, Kyoto University, Kitashirakawa-Oiwakecho, Sakyo-ku, Kyoto 606-8502, Japan.

e-mail: yanagida@kozo.biophys.kyoto-u.ac.jp

1. Marks, J. M. & Hyams, J. S. *Eur. J. Cell Biol.* **39**, 27–32 (1985).
2. Pelham, R. J. Jr & Chang, F. *Nature Cell Biol.* **3**, 235–244 (2001).
3. Gachet, Y., Tournier, S., Millar, J. B. A. & Hyams, J. S. *Nature* **412**, 352–355 (2001).
4. Uhlmann, F. *EMBO Rep.* **2**, 487–492 (2001).
5. Amon, A. *Curr. Opin. Genet. Dev.* **9**, 69–75 (1999).
6. Zachariae, W. & Nasmyth, K. *Genes Dev.* **13**, 2039–2058 (1999).
7. Spector, I., Shochet, N. R., Kashman, Y. & Groweiss, A. *Science* **219**, 493–495 (1983).
8. Wilkinson, M. G. *et al. Genes Dev.* **10**, 2289–2301 (1996).
9. Reinsch, S. & Goczy, P. *J. Cell Sci.* **111**, 2283–2295 (1998).
10. Adler, P. N. & Taylor, J. *Curr. Biol.* **11**, R233–R236 (2001).

Daedalus

Unchemical chemistry

Chemistry is mainly about simple aggregates of just a few atoms. It deals more uneasily with polymers, huge molecules or bonded lattices, but even these redeem themselves by being composed of repeating simple units. Daedalus is now inventing an entirely different chemistry.

His inspiration was the strange yellow solid deposited from phosphine gas. This solid is apparently a random mass of phosphorus atoms bonded any which way, with any loose bonds terminated by hydrogen atoms. He reckons there must be a whole range of random atomic aggregates like this, with no repeating units. Indeed, chemists must have made them thousands of times — as the brown 'tar' of so many failed reactions, always thrown down the sink in disgust. Coal is perhaps the closest natural example. But Daedalus is now making them deliberately.

He is aiming a number of atomic beams *in vacuo* at a cryogenically cooled target. Each atom sticks where it lands, and bonds to other atoms nearby. The overall composition of the product can be varied by changing the intensity of the beams, but its short-range structure must be entirely random. It will not be a polymer or copolymer, but rather a random macromolecular structure, or 'randomer'.

When the deposit is warmed up, it may turn out to be unstable, and decompose to small molecules. But for certain ranges of composition, it will be stable. Stable randomers will probably contain a high proportion of atoms such as carbon, sulphur, silicon and phosphorus, which bond easily to each other. They will have less hydrogen, oxygen, nitrogen or halogens, although these will be useful for terminating spare valencies. Metals, especially transition metals of variable valency, may aid the incorporation of other atoms. They will also add variety to the range and properties of randomers.

When he has a sense of the most promising randomer compositions and their properties, Daedalus will make them in bulk, by adapting the most tar-like products of conventional reactions. It will be slow work, and even Daedalus cannot guess what use its products may be. As covalently bonded glasses, they may fill a gap in the polymer spectrum; as utterly unnatural substances, they may be ideal for inert surgical implants; as solid-state compositions, they may transform electronics. But new chemistry is always fertile. The field of randomers is bound to be good for something.

David Jones

Nile flooding sank two ancient cities

Sediment failure caused these riverbank sites to drown over 1,200 years ago.

Ruins of the ancient cities of Eastern Canopus and Herakleion, dated from Greek to Byzantine times, were recently discovered at depths of 6–7 metres in Egypt's Abu Qir Bay¹, just off the north-western Nile delta, but there is no record of when, how or why they sank. Their submergence has been attributed to the effects of a rise in sea level^{2,3}, subsidence at the delta margin^{2,3} and earthquakes⁴. Here we present new evidence indicating that the disappearance of these sites was primarily the result of riverbank failure in the Canopic region and, in the case of Eastern Canopus, was triggered by flooding of the Nile in the middle of the eighth century AD.

The centres of Eastern Canopus and Herakleion were positioned at the mouth of the Nile's Canopic branch, a fluvial system that served as an important waterway for Greek and subsequent trade^{5,6} until the branch ceased to function late in the first millennium AD^{3,7}. Artefacts have been recovered in recent times by fishermen from the bay, and the sites were first explored by hard-hat divers in 1933 (ref. 2).

Eastern Canopus and Herakleion were discovered during geophysical and diving exploration of an area of over 100 km² in western Abu Qir Bay¹ in 1999 and 2000 (Fig. 1a, b). These sites are 1.6 and 5.4 km, respectively, east of the Abu Qir headland. At each site, ruins were found over an area exceeding 0.5 km² by using bathymetric, side-scan sonar, high-resolution sub-bottom profiling and nuclear-resonance magnetometric surveys, as well as excavation by divers.

These investigations, which are complemented by study of sediment cores collected along the northwestern Nile delta³, revealed that the Canopic branch migrated from east to west and then eastwards again before the development of the Rosetta branch 26 km east of the Abu Qir headland in the second millennium AD. Large features such as walls, bases of temples, columns, stellae and statues, which are apparent on side-scan sonar records, are concentrated on the margins of the now-submerged Canopic channels. Excavation has exposed further structural remains and Ptolemaic Greek, Roman and Byzantine artefacts.

The most recent text pertaining to the sites, written by Sophronius of Damas, Patriarch of Jerusalem, is dated to the early seventh century AD⁶ and records that structures in Eastern Canopus at that time were still positioned close to the shore. The temples and walls must have remained exposed for another century, until after AD 731, because the two youngest coins found

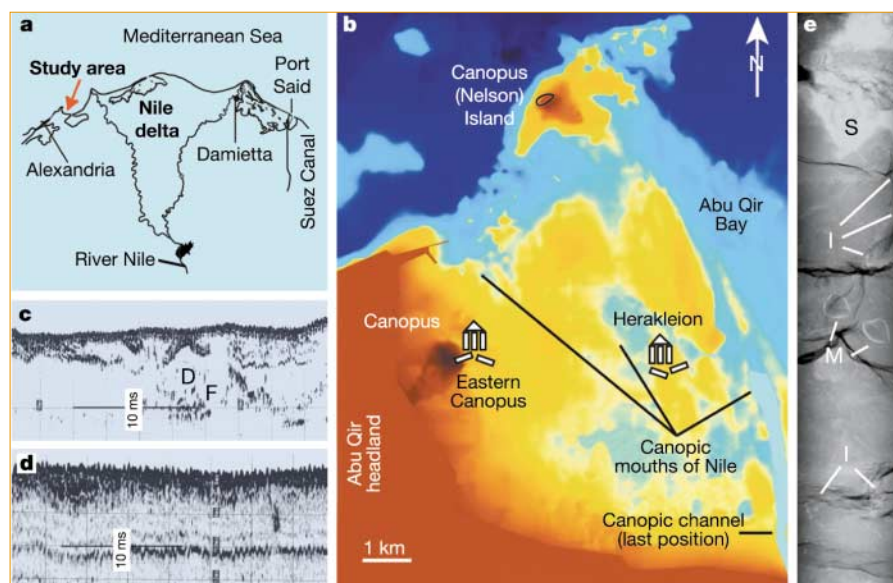


Figure 1 Sediment failure induced destruction of ancient sites and their submergence to below 6 m in waters off Egypt. **a**, Map of the study area. **b**, Eastern Canopus and Herakleion were positioned on the Canopic branch of the Nile, which formerly flowed into western Abu Qir Bay. Over time, the branch channel and mouth migrated east to west and then eastwards again³. Bathymetric depth code: orange, < 4 m; yellow, 4–7 m; light blue, 7–12 m; dark blue, > 12 m. **c**, Seismic profile showing disturbed strata beneath the ruins of Eastern Canopus (D, mud diapir; F, fault offset) restricted to the foundered Canopic channel margin. **d**, In contrast to areas of the channel bank underneath the sites (**c**), most sediment substrate in Abu Qir Bay is undisturbed. Vertical time scale refers to one square. **e**, X-radiographic image of a core section (7 cm in diameter) collected at the top of failed mud substrate directly beneath the ruins of Eastern Canopus and marine sand cover (S). This disturbed stratum comprises discontinuous, non-marine wetland mud into which silt and sand have been vertically injected (I); note also the two hinged-bivalve (*Cardium*) shells (M) that were rapidly buried and entrapped in the failed sediment mass.

at this site (both Arabic) date to around AD 731 and AD 724–743, respectively.

Records of Nile flow at the Roda nilometer in Cairo also provide evidence for this timing. A significant flood occurred in AD 741 or 742, when the Nile rose more than 1 m above the normal high-flood stage⁸. Other floods of this magnitude⁸ are only recorded well before (AD 719 or 720) and long after (AD 805) the AD 741/742 event.

A thin (20–100 cm) shell-rich sand layer covers the sea floor in western Abu Qir Bay, yet mud-rich sediment immediately underneath this modern marine sand comprises ostracods, foraminifera, molluscs and flora of distinctly freshwater to brackish-marsh and lagoon origin, and is comparable to those in present-day wetlands just south of the sites in the Nile delta.

A notable discovery revealed by magnetometry is the presence of several long, curvilinear, V-shaped trenches (roughly 100 × 5 × 2 m) which have basal cracks buried in the mud substrate near the ruins. Recorded on high-resolution seismic profiles, these trenches are crown cracks and are interpreted as surface expressions of growth faults and post-depositional failure of mud-rich sediment near the temples.

Substrate failure began before the final

destruction of the site and its submergence, as indicated by the age of antelope bone (from about 100 AD) found on the trench floor, trench walls artificially lined with marsh plant mats, and trenches filled (also artificially) with beach and dune sand.

Seismic profiles reveal substrate sediment deformed by mud diapiric structure at shallow depths beneath the sites. Fourteen atomic-mass-spectroscopy studies and conventional radiocarbon analysis of plant matter from the mud immediately beneath the ruins give a marked disparity of calibrated dates: nine were of reasonable stratigraphic age, ranging from 520 BC to AD 322, and five were much older (5730 BC to 1310 BC) and were from deeper strata that were uplifted and are now exposed on the sea floor.

Seismic profiles and cores show failed strata in the bay that are specifically restricted to mud-rich margins of the now-submerged Canopic channel (Fig. 1c, d); together with the associated features we document here, this indicates rapid substrate foundering. Offset strata and disrupted mud layers, penetrated by vertically injected sand and silt stringers, are commonly observed in borings (Fig. 1e) and also reflect post-depositional liquefaction, sediment flow and slumping of mud sec-

tions. The submergence must therefore have involved processes other than a progressive rising of sea level and isostatic lowering of land (which together account for less than 3 m of vertical offset since AD 700).

Although wave damage is known to have occurred on the Levantine margin in AD 746 (possibly in AD 743 or 745)⁹, no earthquake activity is recorded in Egypt during this period¹⁰. We attribute Canopic riverbank failure to local weighting by turbulent, sediment-rich Nile waters being suddenly added upon the soft, organic-rich, physically unstable muds. Sudden failure of the low-elevation margins (less than 1 m) of the river banks probably allowed water to flow over them, leading to an eastward lateral shift of the river channel.

Similar processes have occurred at the mouth of the modern Mississippi river in the United States¹¹. Like the Canopic branch of the Nile, this is in a relatively stable area; sediment failure during Mississippi flooding includes liquefaction, slumping on slopes of less than one degree, and diapirism near the river mouths.

Our investigation in Abu Qir Bay indicates that structural failure of the cities that were once positioned on the river banks,

and their submergence to depths of more than 5 m at and near the Canopic mouth, are best explained by sediment failure triggered by flooding of the Nile as recently as 1,250 years ago.

Jean-Daniel Stanley*, Franck Goddio†, Gerard Schnepf†

**Deltas-Global Change Program, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA*

e-mail: stanley.daniel@nmnh.si.edu

†Institut Européen d'Archéologie Sous-Marine, 75 Rue de Grenelle, 70005-Paris, France

1. Goddio, F. *EOS, Trans. Am. Geophys. Union* **81**, F20 (2000).
2. Toussoun, O. *Bull. Soc. R. Archéol. Alexandrie* **29**, 342–352 (1934).
3. Warne, A. G. & Stanley, D. J. *J. Coastal Res.* **9**, 26–64 (1993).
4. Nur, A. M. *EOS, Trans. Am. Geophys. Union* **81**, F21 (2000).
5. Jones, H. L. (transl.) *The Geography of Strabo* (Heinemann, London, 1917).
6. Bernard, A. *Le Delta Egyptien d'Après les Textes Grecs, I. Les Confins Libyques* (Inst. Franç. Arch. Orientale, Cairo, 1970).
7. Toussoun, O. *Mém. Inst. Egypte* **4**, 1–212 (1922).
8. Popper, W. *The Cairo Nilometer* (Univ. California Press, Berkeley, 1951).
9. Saloviev, S. L. et al. *Tsunamis in the Mediterranean Sea 2000 BC–2000 AD* (Kluwer, Dordrecht, 2000).
10. Guidoboni, E. *Catalogue of Ancient Earthquakes in the Mediterranean Area Up to the 10th Century* (Ist. Naz. Geof., Bologna, 1994).
11. Coleman, J. M. *Deltas: Processes of Deposition and Models for Exploration* (Int. Hum. Res. Dev. Corp., Boston, 1982).

Animal behaviour

An unusual social display by gorillas

We have observed wild western lowland gorillas (*Gorilla gorilla gorilla*) using water to generate spectacular 'splash displays'. Most of these displays were made by silverbacks in an agonistic context, and we propose that they are primarily linked to the intimidation of potential rivals for female acquisition. This unusual behaviour may have developed only in gorillas that visit open swampland, where visibility greatly exceeds that encountered in the forest and highly visual, long-distance displays are therefore of value.

Almost nothing is known about the social behaviour of western lowland gorillas because of poor observation conditions and

difficulties of habituation in dense forest. But the discovery that large numbers of gorillas feed in open, swampy clearings (bais) in the forest of northern Congo means that their social interaction can be investigated.

Display behaviour incorporating objects in the environment occurs in agonistic encounters in all ape species¹, but manipulating water for communication has not been described in any wild primate and, with the possible exception of elephants, may be unique among terrestrial mammals.

We observed 124 gorillas over 32 months at the 12.9-hectare 'Mbeli Bai'. Gorillas were visible for 27% (1,681 h) of the time that we were present at the clearing. Ninety splash displays, representing 57 independent bouts of social interaction, were produced by 19 individuals (4 unaffiliated, 'solitary' silverback males and 15 individuals from 9 groups). Ten display styles were seen, of which three were used in 75% of all displays. These were the 'body splash' (35% of displays), in which a gorilla runs or leaps into standing water of up to 1.5 m in depth (Fig. 1), and one-handed and two-handed splashes (40%), in which one or both arms are raised and then brought down forcibly, the open palms striking the water surface at a slight angle. Each of these three techniques generates large plumes of spray.

Of all splash-display bouts, 67% were produced in an agonistic context, 17.5% were made in play and 5% were directed towards other species; in 10.5% of cases the context was not evident. Silverbacks dis-

played the most frequently (68%), and almost exclusively in an agonistic context.

When the observed frequencies of splash bouts were compared with the expected frequencies (calculated from the proportion of visits made to the bai by each age/sex class), group silverbacks displayed twice as often as expected, whereas solitary silverbacks displayed more than four times as often. Solitary silverbacks were also the most frequent recipients of the display (six times more often than expected). Adult females, although they were the most frequently seen age/sex class, never produced splash displays and were rarely targeted.

Directly attracting the attention of females is not considered the prime purpose of splash displays, because solitary males displayed almost as frequently to other solitary males as to groups, and in more than half of these cases no females were in sight. The more likely purpose is to intimidate potential competitors for acquiring females.

Splash displays are an example of object-mediated behavioural plasticity in response to unusual circumstances. Although the three primary display styles resemble the dry-land charges and ground-slap displays seen in many gorilla populations, differences in context, execution and intra-dyad distance confirm that splash displays represent a distinct behavioural element in gorillas' visual-display repertoire.

The bai offers visibility of up to 500 m, which is never encountered by gorillas in the forest, and so long-distance visual displays are clearly expedient. In the wild, the only other record of splash display comes from western lowland gorillas at Maya Bai, 180 km from Mbeli (F. Magliocca, personal communication).

The paucity of data on western lowland gorillas has led to generalizations about their behaviour based on that of mountain gorillas (*G. beringei beringei*). But their feeding ecology is different², and our findings indicate that their social behaviour is too. We anticipate that gorillas, maligned as cognitively poor cousins to the other great apes, will emerge from further bai studies as adaptable, innovative and intelligent creatures that exploit a complex environment³.

Richard J. Parnell*†, Hannah M. Buchanan-Smith*

**Scottish Primate Research Group, Department of Psychology, University of Stirling,*

Stirling FK9 4LA, UK

e-mail: r.j.parnell@stir.ac.uk

†Wildlife Conservation Society, New York

Zoological Society, 185th Street and Southern

Boulevard, Bronx, New York 10460-1099, USA

1. Beck, B. B. *Animal Tool Behavior: The Use and Manufacture of Tools by Animals* (Garland, New York, 1980).
2. Williamson, E. A. et al. *M. Am. J. Primatol.* **21**, 265–277 (1990).
3. Byrne, R. W. in *Reaching into Thought: The Minds of the Great Apes* (eds Russon, A. E., Bard, K. A. & Taylor Parker, S.) 111–130 (Cambridge Univ. Press, Cambridge, 1996).



Figure 1 A solitary silverback performing a 'body splash' display at Mbeli Bai, Nouabalé-Ndoki National Park, Congo (Brazzaville).

Dynamics of collapsing and exploding Bose–Einstein condensates

Elizabeth A. Donley*, Neil R. Claussen*, Simon L. Cornish*, Jacob L. Roberts*, Eric A. Cornell*† & Carl E. Wieman*

* JILA, Campus Box 440, and Department of Physics, Campus Box 390, University of Colorado, Boulder, Colorado 80309, USA

† Quantum Physics Division, National Institute of Standards and Technology, JILA A231, Boulder, Colorado 80309, USA

When atoms in a gas are cooled to extremely low temperatures, they will—under the appropriate conditions—condense into a single quantum-mechanical state known as a Bose–Einstein condensate. In such systems, quantum-mechanical behaviour is evident on a macroscopic scale. Here we explore the dynamics of how a Bose–Einstein condensate collapses and subsequently explodes when the balance of forces governing its size and shape is suddenly altered. A condensate’s equilibrium size and shape is strongly affected by the interatomic interactions. Our ability to induce a collapse by switching the interactions from repulsive to attractive by tuning an externally applied magnetic field yields detailed information on the violent collapse process. We observe anisotropic atom bursts that explode from the condensate, atoms leaving the condensate in undetected forms, spikes appearing in the condensate wavefunction and oscillating remnant condensates that survive the collapse. All these processes have curious dependences on time, on the strength of the interaction and on the number of condensate atoms. Although the system would seem to be simple and well characterized, our measurements reveal many phenomena that challenge theoretical models.

Although the density of an atomic Bose–Einstein condensate (BEC) is typically five orders of magnitude lower than the density of air, the interatomic interactions greatly affect a wide variety of BEC properties. These include static properties, such as the size, shape and stability of the condensate, and dynamic properties, like the collective excitation spectrum and soliton and vortex behaviour. Because all of these properties are sensitive to the interatomic interactions, they can be dramatically affected by tuning the strength and sign of the interactions.

Most of the physical processes in BECs are well described by mean-field theory¹, in which the strength of the interactions depends on the atom density and on one additional parameter called the *s*-wave scattering length, *a*. The parameter *a* is determined by the atomic species. When *a* > 0, the interactions are repulsive. In contrast, when *a* < 0, the interactions are attractive and a BEC tends to contract to minimize its overall energy. In a magnetic trap, the contraction competes with the kinetic zero-point energy, which tends to spread out the condensate. For a strong enough attractive interaction, there is not enough kinetic energy to stabilize the BEC, and it is expected to implode. A BEC can avoid implosion only as long as the number of atoms in it, *N*₀, is less than a critical value given by²

$$N_{\text{cr}} = ka_{\text{ho}}/|a| \quad (1)$$

where the dimensionless constant *k* is called the stability coefficient. The precise value of *k* depends on the ratio of the magnetic trap frequencies³. *a*_{ho} is the harmonic oscillator length, which sets the size of the condensate in the ideal-gas (*a* = 0) limit.

Under most circumstances, *a* is insensitive to external fields. But in the vicinity of a so-called Feshbach resonance, *a* can be tuned over a very large range by adjusting the externally applied magnetic field^{4,5}. This has been demonstrated in recent years with cold ⁸⁵Rb and Cs atoms^{6–8}, and with BECs of Na and ⁸⁵Rb (refs 9, 10). For ⁸⁵Rb atoms, *a* is usually negative, but a Feshbach resonance at ~155 G allows us to tune *a* by orders of magnitude and even change its sign. This gives us the ability to create stable BECs of ⁸⁵Rb (ref. 10) and to adjust the interatomic interactions. We recently used this flexibility to verify the functional form of equation (1), and to measure the stability coefficient to be *k* = 0.46(6) (ref. 11).

Here we study the dynamical response (‘the collapse’) of an initially stable BEC to a sudden shift of the scattering length to a value more negative than the critical value *a*_{cr} = −*ka*_{ho}/*N*₀. We

have observed many features of the surprisingly complex collapse process, including the energies and energy anisotropies of atoms that burst from the condensate, the timescale for the onset of this burst, the rates for losing atoms, spikes in the wavefunction that form during collapse, and the size of the remnant BEC that survives the collapse. The level of control provided by tuning *a* has allowed us to see how all of these quantities depend on the magnitude of *a*, on the initial number and density of condensate atoms, and on the initial spatial size and shape of the BEC before the transition to instability.

A great deal of theoretical interest^{12–17} was generated by BEC experiments in ⁷Li (ref. 18), for which the scattering length is also negative and collapse events are also observed^{19,20}. The ⁷Li experiments do not use a Feshbach resonance, so *a* is fixed. This restricts experimentation to the regime where the initial number of condensate atoms is less than or equal to *N*_{cr}, and the collapse is driven by a stochastic process. In addition, studies of collapse dynamics in ⁷Li are complicated by a large thermal component. Our ability to tune the scattering length, and to explore the regime where the initial condensate is ‘pure’ (near *T* = 0) and the number *N*₀ is much larger than *N*_{cr}, allows us to explore the dynamics and compare it with theory in far more detail.

Experimental techniques

The procedure for producing stable ⁸⁵Rb condensates has been described in detail elsewhere¹⁰. A standard double magneto-optical trap (MOT) system²¹ was used to collect a cold sample of ⁸⁵Rb atoms in a low-pressure chamber. Once sufficient atoms had accumulated in the low-pressure MOT, the atoms were loaded into a cylindrically symmetric cigar-shaped magnetic trap with frequencies *ν*_{radial} = 17.5 Hz and *ν*_{axial} = 6.8 Hz. Radio-frequency evaporation was then used to cool the sample to ~3 nK to form pure condensates containing >90% of the sample atoms. The final stages of evaporation were performed at 162 G, where the scattering length is positive and stable condensates of up to 15,000 atoms could be formed. After evaporative cooling, the magnetic field was increased adiabatically to 166 G (except where noted), where *a* = 0. This provided a well-defined initial condition, with the BEC taking on the size and shape of the harmonic oscillator ground state.

We could then adjust the mean-field interactions within the BEC to a variety of values on timescales as short as 0.1 ms. The obvious manipulation was to jump to some value of *a* < *a*_{cr} to trigger a

collapse, but the tunability of a was also a great help in imaging the sample. Usually the condensate size was below the resolution limit of our imaging system ($7\text{ }\mu\text{m}$ full-width at half-maximum). However, we could increase the scattering length to large positive values and use the repulsive interatomic interactions to expand the BEC before imaging, thus obtaining information on the pre-expansion condensate shape and number. A typical $a(t)$ sequence is shown in Fig. 1a. We have used a variety of such sequences to explore many aspects of the collapse, and enhance the visibility of particular components of the sample.

Condensate contraction and atom loss

When the scattering length is changed rapidly to a value $a_{\text{collapse}} < a_{\text{cr}}$, a condensate's kinetic energy no longer provides a sufficient barrier against collapse. As described in ref. 13, during collapse one might expect a BEC to contract until losses from density-dependent inelastic collisions²² would effectively stop the contraction. This contraction would take place roughly on the timescale of a trap oscillation, and the density would sharply increase after $T_{\text{rad}}/4 \approx 14\text{ ms}$, where T_{rad} is the radial trap period.

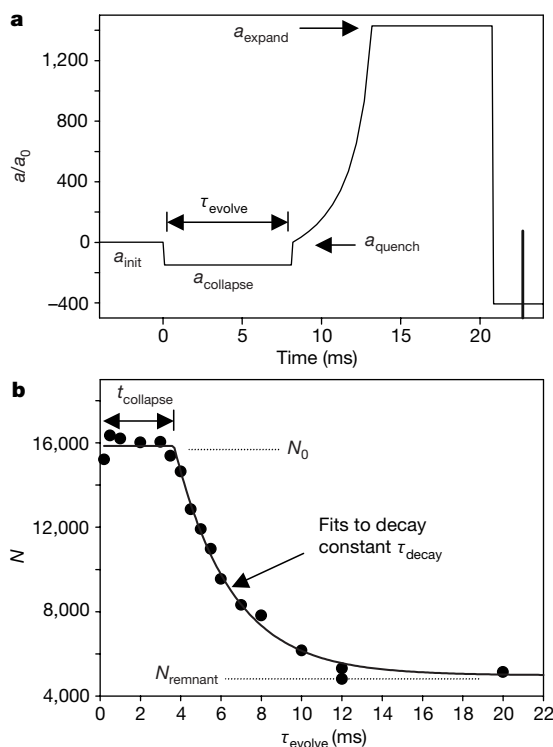


Figure 1 An example of a ramp applied to the scattering length a , and a plot of the condensate number N versus time after a jump to a negative scattering length. **a**, A typical $a(t)$ sequence. $a_0 = 0.529\text{ }\text{\AA}$ is the Bohr radius. The scattering length is jumped at $t = 0$ in 0.1 ms from a_{init} to a_{collapse} , where the BEC evolves for a time τ_{evolve} . The field is carefully controlled so that magnetic-field noise translates into fluctuations in a_{collapse} on the order of $\sim 0.1a_0$ in magnitude. The collapse is then interrupted with a jump to a_{quench} , and the field is ramped in 5 ms to a large positive scattering length which makes the BEC expand. After 7.5 ms of additional expansion, the trap is turned off in 0.1 ms , and 1.8 ms later the density distribution is probed using destructive absorption imaging with a $40\text{-}\mu\text{s}$ laser pulse (indicated by the vertical bar). The increase in a from a_{collapse} to a_{expand} is far too rapid to allow for the BEC to expand adiabatically. On the contrary, the smaller the BEC before expansion, the larger the cloud at the moment of imaging. Thus we can readily infer the relative size of the bulk of the BEC just before the jump to a_{quench} . The density of the expanded BEC is so low that the rapid transit of the Feshbach resonance pole²⁵ during the trap turn-off and the subsequent time spent at magnetic field $B = 0$ ($a = -400a_0$) both have a negligible effect. **b**, The number of atoms remaining in the BEC versus τ_{evolve} at $a_{\text{collapse}} = -30a_0$. We observed a delayed and abrupt onset of loss. The solid line is a fit to an exponential with a best-fit value of $t_{\text{collapse}} \approx 3.7(5)\text{ ms}$ for the delay.

How does this picture compare to what we have seen?

A plot of the number of atoms in the condensate, N , versus τ_{evolve} for $a_{\text{collapse}} = -30a_0$ and $a_{\text{init}} = +7a_0$ is shown in Fig. 1b. The number remained constant for some time after the jump in a until atom loss suddenly began at t_{collapse} . After the jump the condensate was smaller than our resolution limit, so we could not observe the contraction directly, but we could infer the extent of the contraction by observing the degree of mean-field expansion. We observed that the post-expansion condensate widths changed very little with time τ_{evolve} before t_{collapse} . From this we infer that the bulk BEC did not contract dramatically before loss began. The equations in ref. 23 (in the gaussian approximation) predict fairly well the observed expanded shape over a fairly large range of a_{collapse} and t before collapse. So we used these equations to estimate the density before collapse, and find that the predicted contraction only corresponds to a 50% increase in the average density to $2.5 \times 10^{13}\text{ cm}^{-3}$. Using the decay constants from ref. 22, this density gives an atom loss rate, τ_{decay} , that is far smaller than what we observe and does not have the observed sudden onset.

For the data in Fig. 1b and most other data presented below, a was changed (in 0.1 ms) to $a_{\text{quench}} = 0$ after a time τ_{evolve} at a_{collapse} . We believe that the loss stopped immediately after this jump in a . This interpretation is based on the observation that the quantitative details of curves such as that shown in Fig. 1b did not depend on whether the collapse was terminated by a jump to $a_{\text{quench}} = 0$ or to $a_{\text{quench}} = 250a_0$.

We have measured loss curves like that in Fig. 1b for many different values of a_{collapse} . The collapse time versus a_{collapse} for $N_0 = 6,000$ presented in Fig. 2 shows a strong dependence on a_{collapse} . Reducing the initial density by a factor of ~ 4 (with a corresponding increase in volume) by setting $a_{\text{init}} = +89a_0$ for one value of a_{collapse} ($-15a_0$), increased t_{collapse} by a factor of about three.

The atom loss time constant τ_{decay} depended only weakly on a_{collapse} and N_0 . For the range of a_{collapse} shown in Fig. 2, τ_{decay} did not depend on a_{collapse} or N_0 outside of the experimental noise ($\sim 20\%$). On average, τ_{decay} was $2.8(1)\text{ ms}$. For the very negative value of $a_{\text{collapse}} \approx -250a_0$, however, τ_{decay} did decrease to 1.8 ms for $N_0 = 6,000$ and to 1.2 ms for $N_0 = 15,000$.

Bursts of atoms. Atoms leave the BEC during the collapse (Fig. 1b). There are at least two components to the expelled atoms. One component (the ‘missing atoms’) is not detected. The other component emerges as a burst of detectable, spin-polarized atoms with energies much greater than the initial energy of the condensate but much less than the magnetic trap depth. The dependences of burst energy on a_{collapse} and N_0 are complex, but as they will provide a stringent test of collapse theories, we present them in detail.

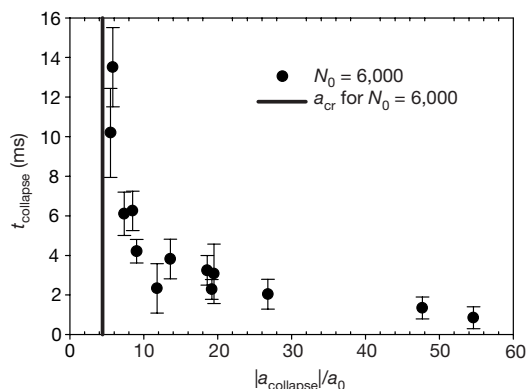


Figure 2 The collapse time t_{collapse} versus the scattering length to which the condensate is jumped, a_{collapse} for $6,000$ -atom condensates. The vertical line indicates the critical value of the scattering length, a_{cr} for $N_0 = 6,000$. The data were acquired with $a_{\text{init}} = a_{\text{quench}} = 0$ (to within $\sim 2a_0$; see Fig. 1).

The angular kinetic-energy distribution with which the burst atoms are expelled from the condensate can most accurately be measured by observing their harmonic oscillations in the trap (Fig. 3a). For example, half of a radial period after the expulsion ($T_{\text{rad}}/2$), all atoms return to their initial radial positions. Well before or well after this ‘radial focus’, the burst cloud is too dilute to be observed. Fortunately, at the radial focus, oscillations along the axial trap axis are near their outer turning points, and the axial energy can be found from the length of the ‘stripe’ of atoms along the axial axis. The radial energy can be found with the same procedure for an axial focus. The sharpness of the focus also provides information on the time extent of the burst.

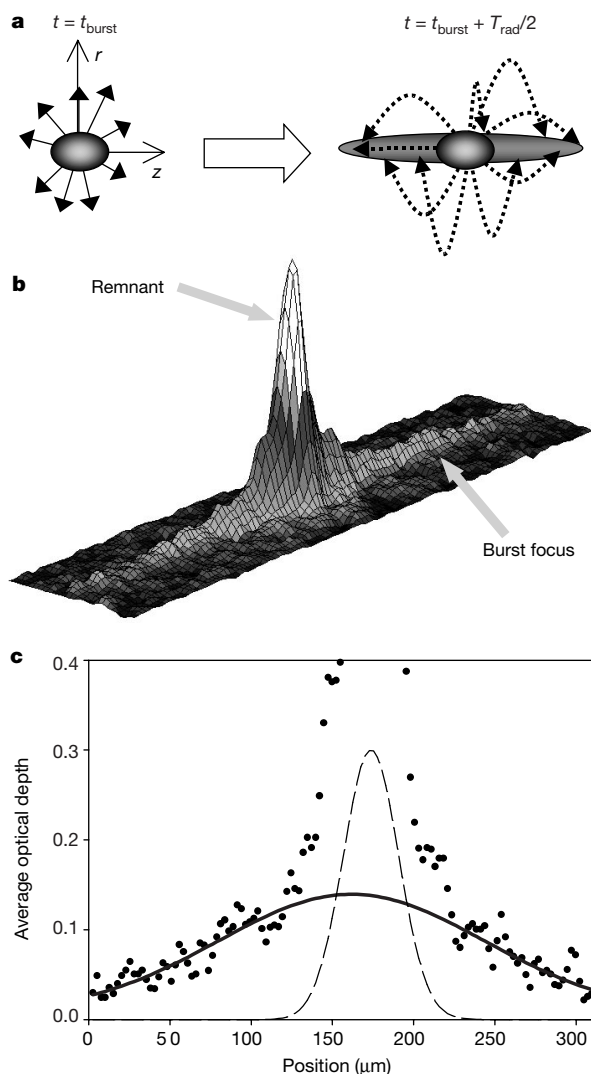


Figure 3 A burst focus. **a**, Conceptual illustration of a radial burst focus. t_{burst} is the time at which the burst is generated and T_{rad} is the radial trap period. **b**, An image of a radial burst focus taken 33.5 ms after a jump from $a_{\text{init}} = 0$ to $-30a_0$ for $N_0 = 10,000$. $T_{\text{rad}}/2 = 28.6$ ms, which indicates that the burst occurred 4.9(5) ms after the jump. The axial energy distribution for this burst corresponded to an effective temperature of 62 nK. The image is $60 \times 310 \mu\text{m}$. **c**, Radially averaged cross-section of **b** with a gaussian fit to the burst energy distribution. The central $100 \mu\text{m}$ were excluded from the fit to avoid distortion in the fit due to the condensate remnant ($\sigma = 9 \mu\text{m}$) and the thermal cloud ($\sigma = 17 \mu\text{m}$). The latter is present in the pre-collapse sample due to the finite temperature, and appears to be unaffected by the collapse. The dashed line indicates the fit to this initial thermal component. We note the offset between the centres of the burst and the remnant. This offset varies from shot to shot by an amount comparable to the offset shown.

Figure 3b shows an image of a radial focus. The size scales for the burst focus and the remnant were well separated because the latter was not expanded before imaging. Figure 3c shows cross-sections of the burst focus, and fits to the burst and to the thermal cloud. The burst energy distributions were well fitted by gaussians characterized by a temperature that was usually different for the two trap directions. The burst energy fluctuated from shot to shot by up to a factor of 2 for a given a_{collapse} . This variation is far larger than the measurement uncertainty or the variation in initial number (both $\sim 10\%$), and its source is unknown. (We also discuss observed structural variability when we present the jet measurements below.) Although the burst energies varied from shot to shot, the average

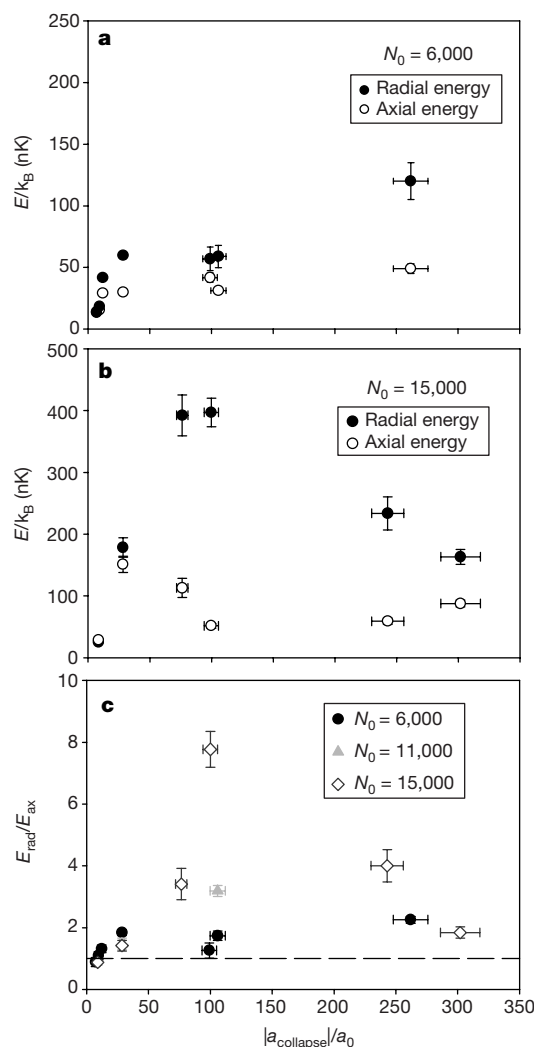


Figure 4 Burst energies and energy anisotropies. **a** and **b**, the axial and radial burst energies versus a_{collapse} for $N_0 \approx 6,000$ and $N_0 \approx 15,000$, respectively. On average, ten burst focuses were measured for each trap direction at each value of a_{collapse} studied. The energies were higher for the larger- N_0 condensates over the full range of a_{collapse} studied. The vertical and horizontal error bars indicate respectively the standard error of the measurements and the uncertainty in a_{collapse} arising from the magnetic-field calibration. For several of the points, the uncertainties are smaller than the symbol size and the error bars are not visible. **c**, The ratio of the radial to the axial energies, $E_{\text{rad}}/E_{\text{ax}}$, which is a measure of the burst anisotropy. For values of $|a_{\text{collapse}}|$ just past a_{cr} , the burst was isotropic for both $N_0 \approx 6,000$ and $N_0 \approx 15,000$. At larger values of $|a_{\text{collapse}}|$, larger- N_0 condensates gave rise to stronger anisotropies. For $N_0 = 6,000$, a_{init}/a_0 was 0 and λ was 1.6, and for $N_0 = 15,000$, a_{init}/a_0 was +7 and λ was 2, where $\lambda = \sigma_z/\sigma_r$ is the aspect ratio of the initial condensate. When instead we started at $a_{\text{init}} = +100a_0$, the BEC was initially more anisotropic ($\lambda = 2.4$), but the burst became more isotropic, with E_{ax} going up by $\sim 40\%$ and E_{rad} dropping by $\sim 60\%$ for $N_0 = 15,000$ and $a_{\text{collapse}} = -100a_0$.

value was well-defined and showed trends far larger than the variation. The axial and radial burst energies versus a_{collapse} are shown in Fig. 4a and b for $N_0 = 6,000$ and $15,000$, respectively. The burst-energy anisotropy shown in Fig. 4c depended on N_0 , a_{collapse} and a_{init} .

When we interrupted the collapse with a jump back to $a = 0$ as discussed above, we also interrupted the growth of the burst. The ‘interrupted’ burst atoms still refocused after sitting at $a = 0$ for the requisite half trap period. The energy of the atoms in the interrupted bursts appeared to be the same, but the number of atoms was smaller. By changing the time at which the collapse was interrupted, we could measure the time dependence of the creation of burst atoms. For the conditions of Fig. 1b, the number of burst atoms N_{burst} grew with τ_{evolve} with a time constant of 1.2 ms, starting at 3.5 ms and reaching an asymptotic final number of $\sim 2,500$ for all times ≥ 7 ms. N_{burst} varied randomly by $\sim 20\%$ for the data in Fig. 4, but on average the fraction of atoms going into the burst was about 20% of N_0 and did not depend on a_{collapse} or N_0 .

Remnant condensate. After a collapse, a ‘remnant’ condensate containing a fraction of the atoms survived with nearly constant number for more than 1 second, and oscillated in a highly excited collective state with the two lowest modes ($\nu \approx 2\nu_{\text{axial}}$ and $\nu \approx 2\nu_{\text{radial}}$) being predominantly excited. (The measured frequencies were $\nu = 13.6(6)$ Hz and $\nu = 33.4(3)$ Hz.) To find the oscillation frequencies, the widths of the condensate were measured as a function of time spent at a_{collapse} .

The number of atoms in the remnant depended on a_{collapse} and N_0 , and in general was not limited by the critical number, N_{cr} . The stability condition in equation (1) determined the collapse point, but did not constrain N_{remnant} . A fixed fraction of N_0 went into the remnant independent of N_0 , so that smaller condensates often ended up with $N_{\text{remnant}} < N_{\text{cr}}$, but larger condensates rarely did. The fraction of atoms that went into the remnant decreased with $|a_{\text{collapse}}|$, and was $\sim 40\%$ for $|a_{\text{collapse}}| < 10a_0$ and $\sim 10\%$ for $|a_{\text{collapse}}| > 100a_0$. We do not think that surface-wave excitations²⁴ are responsible for stabilizing the remnant because we excite large-amplitude breathing modes. For $N_0 = 6,000$ and $|a_{\text{collapse}}| < 10a_0$, more atoms were lost than the number required to lower N_{remnant} to below N_{cr} .

As N_{burst} was independent of a_{collapse} but N_{remnant} decreased with $|a_{\text{collapse}}|$, the number of missing atoms increased with $|a_{\text{collapse}}|$. The number of missing atoms also increased with N_0 , but the percentage of missing atoms was equal for $N_0 = 6,000$ and $N_0 = 15,000$ with a_{collapse} fixed and was $\sim 40\%$ for $|a_{\text{collapse}}| < 10a_0$ and $\sim 70\%$ for $|a_{\text{collapse}}| \geq 100a_0$. The missing atoms were presumably either expelled from the condensate at such high energies that we could not detect them ($> 20 \mu\text{K}$), or they were transferred to untrapped atomic states or undetected molecules.

Jet formation. Under very specific experimental conditions, we observed streams of atoms with highly anisotropic velocities emerging from the collapsing condensate. These ‘jets’ are distinguished from the ‘burst’ in that the jets have much lower kinetic energy (on the order of a few nanokelvin), in that their velocity is nearly purely radial, and in that they appear only when the collapse is interrupted (that is, by jumping to $a_{\text{quench}} = 0$) during the period of number loss. Collapse processes that were allowed to evolve to completion (until $N \approx N_{\text{remnant}}$) were not observed to emit jets. Examples are shown in Fig. 5 for different τ_{evolve} for the conditions of Fig. 1b. The size and shape of the jets varied from image to image even when all conditions were unchanged, and as many as three jets were sometimes observed to be emitted from the collapse of a single condensate. Also, the jets were not always symmetric about the condensate axis.

We believe that these jets are manifestations of local ‘spikes’ in the condensate density that form during the collapse and expand when the balance of forces is changed by quenching the collapse. We can estimate the size of the spikes using the uncertainty principle. After a

jump to $a_{\text{quench}} = 0$, the kinetic energy per atom in the resulting jet is equal to the confinement energy that the spike had before quenching the collapse, that is, $(1/2)mv^2 = \hbar^2/(4m\sigma^2)$, where σ is the width of the spike in the wavefunction, m is the mass, and v is the expansion velocity. The anisotropy of the jets indicates that the spikes from which they originated were also highly anisotropic, being narrower in the radial direction. From the widths and the number of atoms in the jets, we can estimate the density in the spikes. Plots of the number of jet atoms and the inferred density in the spikes versus τ_{evolve} are presented in Fig. 6. The jets exhibited variability in energy and number that was larger than the $\sim 10\%$ measurement noise.

Overview of the current theoretical models

Several theoretical papers^{12–17} have considered the problem of collapse of a BEC when the number of atoms exceeds the critical number. These treatments all use a mean-field approach, and describe the condensate dynamics using the Gross–Pitaevskii (GP) equation. In most cases^{12,13,15,16}, the loss mechanism is three-body recombination, but Duine and Stoof¹⁷ propose that the loss arises from a new elastic scattering process. In both cases the loss is density dependent, and so the loss rate is quite sensitive to the dynamics of the shape of the condensate. Because a full three-dimensional anisotropic time-dependent solution to the GP equation is very difficult, these calculations have used various approximations to calculate the time evolution of the condensate shape. Kagan, Muryshev and Shlyapnikov¹³ numerically integrate the GP equation for the case of an isotropic trap and large values of the three-body recombination coefficient K_3 . In this regime, the condensate smoothly contracts in a single, collective collapse. Saito and Ueda^{15,16} perform a similar numerical solution for the isotropic case but with smaller values of K_3 , and observe localized spikes to form in the wavefunction during collapse. Duine and Stoof¹⁷ model the dynamics for the anisotropic case, but use a gaussian approximation rather than an exact numerical solution.

These calculations have all been performed over a certain range of parameters, but none have been done for the specific range of parameters that correspond to our experimental situations. None of the predictions in these papers match our measurements except for

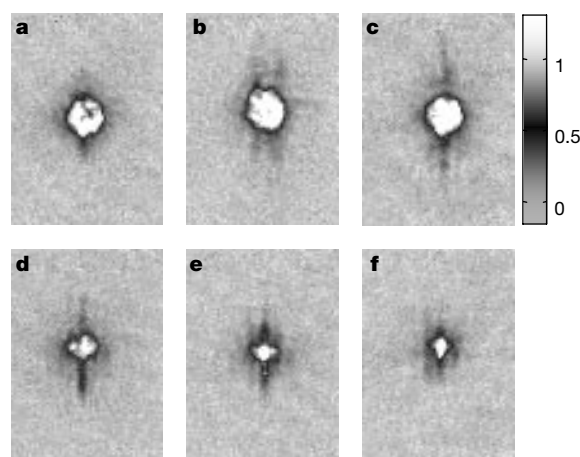


Figure 5 Jet images for a series of τ_{evolve} values for the conditions of Fig. 1b. The evolution times were 2, 3, 4, 6, 8 and 10 ms (from **a** to **f**). Each image is $150 \times 255 \mu\text{m}$. The bar indicates the optical depth scale. An expansion to $a_{\text{expand}} = +250a_0$ was applied, so the jets are longer than for the quantitative measurements explained in the text. The jets were longest (that is, most energetic) and contained the most atoms at values of τ_{evolve} for which the slope of the loss curve (Fig. 1b) was greatest. A tiny jet is barely visible for $\tau_{\text{evolve}} \approx 2$ ms (**a**), which is 1.7 ms before t_{collapse} . The images also show how the number of condensate atoms decreases with time. The time from the application of a_{quench} until the acquisition of the images was fixed at 5.2 ms.

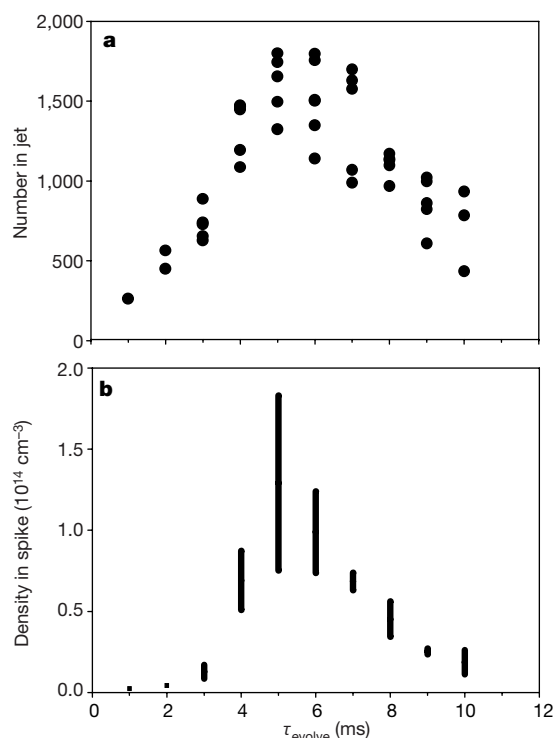


Figure 6 Quantitative jet measurements. **a**, The number of atoms in the jets versus τ_{evolve} for the conditions of Fig. 1b. **b**, The spike density inferred from the kinetic energies of the jets. The bars indicate the range of shot-to-shot variability. For the analysis, we assumed the jets were disk-shaped, as the magnetic trap is axially symmetric. The images were taken perpendicular to the axial trap axis, viewing the disks edge-on. The jets expanded with velocity $v \approx 1 \text{ mm s}^{-1}$, which corresponds to a kinetic energy of $\sim 6 \text{ nK}$ and a radial pre-quench gaussian r.m.s. width of $\sim 0.5 \mu\text{m}$. As the axial size was below our resolution limit, we could not measure the axial expansion rate. For estimating the spike density, we assumed an axial width equal to the harmonic oscillator length. The atom density in the spikes decreased for larger values of $|a_{\text{collapse}}|$, and was half as large for $a_{\text{collapse}} = -100a_0$ as for $-30a_0$.

the general feature that atoms are lost from the condensate. Also, we see several phenomena that are not discussed in these papers. Whether this lack of agreement is due to the fact that these calculations do not scale to our experimental situation or do not contain the proper physics remains to be seen.

Theoretical challenges. Collapsing ^{85}Rb condensates are simple systems with dramatic behaviour. This behaviour might provide a rigorous test of mean-field theory when it is applied to our experimental conditions. Some of our particularly puzzling results are as follows. (1) The decay constant τ_{decay} is independent of both N_0 and $|a_{\text{collapse}}|$ for $|a_{\text{collapse}}| < 100a_0$, and only weakly depends on these quantities for larger $|a_{\text{collapse}}|$. (2) The burst energy per atom dramatically increases with initial condensate number. (3) The fraction of burst atoms is constant when a_{collapse} changes. (4) The number of cold remnant BEC atoms surviving the collapse varies between much less and much more than N_{cr} , depending on N_0 and a_{collapse} but the fractions of remnant atoms, burst atoms, and missing atoms are independent of N_0 .

Outlook

From the experimental point of view, there are at least two questions to be answered. First, is the burst coherent? It may be possible to answer this by generating a sequence of ‘half bursts’ and see if they interfere. Second, where do the missing atoms go? If molecules and/or relatively high-energy atoms are being created, can we detect them? It is clear that adjustable interactions open up new avenues for BEC studies. □

Received 23 April; accepted 18 June 2001.

- Dalfovo, F., Giorgini, S., Pitaevskii, L. P. & Stringari, S. Theory of Bose-Einstein condensation in trapped gases. *Rev. Mod. Phys.* **71**, 463–512 (1999).
- Ruprecht, P. A., Holland, M. J., Burnett, K. & Edwards, M. Time-dependent solution of the nonlinear Schrödinger equation for Bose-condensed trapped neutral atoms. *Phys. Rev. A* **51**, 4704–4711 (1995).
- Gammal, A., Frederico, T. & Tomio, L. Critical number of atoms for attractive Bose-Einstein condensates with cylindrically symmetrical traps. *Phys. Rev. A* (submitted).
- Tiesinga, E., Morderdijk, A. J., Verhaar, B. J. & Stoof, H. T. C. Conditions for Bose-Einstein condensation in magnetically trapped atomic cesium. *Phys. Rev. A* **46**, R1167–R1170 (1992).
- Tiesinga, E., Verhaar, B. J. & Stoof, H. T. C. Threshold and resonance phenomena in ultracold ground-state collisions. *Phys. Rev. A* **47**, 4114–4122 (1993).
- Courteille, Ph., Freeland, R. S., Heinzen, D. J., van Abeelen, F. A. & Verhaar, B. J. Observation of a Feshbach resonance in cold atom scattering. *Phys. Rev. Lett.* **81**, 69–72 (1998).
- Roberts, J. L. *et al.* Resonant magnetic field control of elastic scattering in cold ^{85}Rb . *Phys. Rev. Lett.* **81**, 5109–5112 (1998).
- Vuletić, V., Kerman, A. J., Chin, C. & Chu, S. Observation of low-field Feshbach resonances in collisions of cesium atoms. *Phys. Rev. Lett.* **82**, 1406–1409 (1999).
- Inouye, S. *et al.* Observation of Feshbach resonances in a Bose-Einstein condensate. *Nature* **392**, 151–154 (1998).
- Cornish, S. L., Claussen, N. R., Roberts, J. L., Cornell, E. A. & Wieman, C. E. Stable ^{85}Rb Bose-Einstein condensates with widely tunable interactions. *Phys. Rev. Lett.* **85**, 1795–1798 (2000).
- Roberts, J. L. *et al.* Controlled collapse of a Bose-Einstein condensate. *Phys. Rev. Lett.* **86**, 4211–4214 (2001).
- Kagan, Y., Sukrov, E. L. & Shlyapnikov, G. V. Evolution and global collapse of trapped Bose condensates under variations of the scattering length. *Phys. Rev. Lett.* **79**, 2604–2607 (1997).
- Kagan, Y., Muryshev, A. E. & Shlyapnikov, G. V. Collapse and Bose-Einstein condensation in a trapped Bose gas with negative scattering length. *Phys. Rev. Lett.* **81**, 933–937 (1998).
- Ueda, M. & Huang, K. Fate of a Bose-Einstein condensate with an attractive interaction. *Phys. Rev. A* **60**, 3317–3320 (1999).
- Saito, H. & Ueda, M. Power laws and collapsing dynamics of a trapped Bose-Einstein condensate with attractive interactions. *Phys. Rev. A* **63**, 043601-8 (2001).
- Saito, H. & Ueda, M. Intermittent implosion and pattern formation of trapped Bose-Einstein condensates under attractive interaction. *Phys. Rev. Lett.* **86**, 1406–1409 (2001).
- Duine, R. A. & Stoof, H. T. C. Explosion of a collapsing Bose-Einstein condensate. *Phys. Rev. Lett.* **86**, 2204–2207 (2001).
- Bradley, C. C., Sackett, C. A. & Hulet, R. G. Bose-Einstein condensation of lithium: Observation of limited condensate number. *Phys. Rev. Lett.* **78**, 985–989 (1997).
- Sackett, C. A., Gerton, J. M., Welling, M. & Hulet, R. G. Measurements of collective collapse in a Bose-Einstein condensate with attractive interactions. *Phys. Rev. Lett.* **82**, 876–879 (1999).
- Gerton, J. M., Strekalov, D., Prodan, I. & Hulet, R. G. Direct observation of growth and collapse of a Bose-Einstein condensate with attractive interactions. *Nature* **408**, 692–695 (2000).
- Myatt, C. J., Newbury, N. R., Ghrist, R. W., Loutzenhiser, S. & Wieman, C. E. Multiply loaded magneto-optical trap. *Opt. Lett.* **21**, 290–292 (1996).
- Roberts, J. L., Claussen, N. R., Cornish, S. L. & Wieman, C. E. Magnetic field dependence of ultracold inelastic collisions near a Feshbach resonance. *Phys. Rev. Lett.* **85**, 728–731 (2000).
- Pérez-García, V. M., Michinel, H., Cirac, J. L., Lewenstein, M. & Zoller, P. Dynamics of Bose-Einstein condensates: Variational solutions of the Gross-Pitaevskii equations. *Phys. Rev. A* **56**, 1424–1432 (1997).
- Pattanayak, A. K., Gammal, A., Sackett, C. A. & Hulet, R. G. Stabilizing an attractive Bose-Einstein condensate by driving a surface collective mode. *Phys. Rev. A* **63**, 033604-4 (2001).
- Stenger, J. *et al.* Strongly enhanced inelastic collisions in a Bose-Einstein condensate near Feshbach resonances. *Phys. Rev. Lett.* **82**, 2422–2425 (1999).

Acknowledgements

We thank S. Thompson for laboratory assistance, and S. Dürr, G. Shlyapnikov, H. Stoof, M. Holland, M. Ueda and R. Duine for discussions. This work was supported by the ONR, NSF, ARO-MURI and NIST. S.L.C. acknowledges the support of a Lindemann Fellowship.

Correspondence and requests for materials should be addressed to E.A.D. (e-mail: donley@jilau1.colorado.edu).

Plasma cell differentiation requires the transcription factor XBP-1

Andreas M. Reimold^{*†‡§}, Neal N. Iwakoshi^{*†}, John Manis^{||}, Prashanth Vallabhajosyula^{*}, Eva Szomolanyi-Tsuda[¶], Ellen M. Gravalles[#], Daniel Friend[‡], Michael J. Grusby^{*‡}, Frederick Alt^{||} & Laurie H. Glimcher^{*‡}

^{*} Department of Immunology and Infectious Diseases, Harvard School of Public Health; [‡] Department of Medicine, Harvard Medical School; ^{||} Howard Hughes Medical Institute, Children's Hospital; and [#] Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02115, USA

[¶] Department of Pathology, University of Massachusetts Medical School, Worcester, Massachusetts 01655, USA

[†] These authors contributed equally to this work.

Considerable progress has been made in identifying the transcription factors involved in the early specification of the B-lymphocyte lineage. However, little is known about factors that control the transition of mature activated B cells to antibody-secreting plasma cells. Here we report that the transcription factor XBP-1 is required for the generation of plasma cells. XBP-1 transcripts were rapidly upregulated *in vitro* by stimuli that induce plasma-cell differentiation, and were found at high levels in plasma cells from rheumatoid synovium. When introduced into B-lineage cells, XBP-1 initiated plasma-cell differentiation. Mouse lymphoid chimaeras deficient in XBP-1 possessed normal numbers of activated B lymphocytes that proliferated, secreted cytokines and formed normal germinal centres. However, they secreted very little immunoglobulin of any isotype and failed to control infection with the B-cell-dependent polyoma virus, because plasma cells were markedly absent. XBP-1 is the only transcription factor known to be selectively and specifically required for the terminal differentiation of B lymphocytes to plasma cells.

Production of antibodies by B lymphocytes is critical for the successful removal of pathogens. Once B lymphocytes encounter antigens, a complex series of activation and maturation events drives the generation of memory B cells and antibody-secreting plasma cells^{1–3}. The initial activation of B cells is intimately linked to the ligation of cell-surface receptors and stimulation by cytokines. CD40 is a vital cell-surface molecule whose engagement on B cells results in proliferation, cell survival and the preferential formation of memory cells rather than plasma cells⁴. In addition, cytokines play an essential role in providing signals that influence cell proliferation (for example, interleukin-2 (IL-2), IL-5 and IL-6), protection from apoptosis (IL-4), choice of immunoglobulin isotype during class switching (for example, IL-4 enhances production of IgE), and choice of memory- or plasma-cell phenotype (IL-6 and IL-10)^{3,5–13}. From such stimuli, IgM-secreting plasma cells are generated by T-cell-independent B-cell responses as well as B-cell activation in the extrafollicular zones of lymph nodes, whereas memory cells and non-IgM-secreting plasma cells result from immunoglobulin class switching and somatic hypermutation in germinal centres of lymph nodes¹⁰.

Transcription factors ultimately integrate the signals from cytokine receptors and other activating cell-surface molecules to play essential roles in B-cell maturation and activation¹⁴. When a B cell encounters antigen, the initial signalling events depend on the actions of transcription factors such as BSAP (B-cell-specific activator protein), NF- κ B/Rel, Aiolos and the Ets family members PU.1 and Spi-B^{14,15}. In contrast to these early steps in B-cell development and activation, little is known about the signals that control the transition from an activated B cell to a plasma cell. Plasma cells specialize in the synthesis and secretion of immunoglobulin, and they downregulate a large number of cell-surface molecules (B cell receptor, major histocompatibility complex (MHC) class II, CD20, CD44, B220) and transcription factors (CIITA, BSAP/Pax-5)^{16,17} that are not vital for the terminally differentiated phenotype. Among the few markers that distinguish plasma cells from mature

B cells are cell-surface Syndecan-1, as well as expression of transcription factors IRF-4 (interferon regulatory factor-4) and Blimp-1 (B-lymphocyte-induced maturation protein-1). To date, the only transcription factor known to drive B-cell differentiation to plasma cells is Blimp-1 (also called PRDF1-BF1)^{18–20}. This zinc-finger protein is specifically expressed in mature B cells and plasma cells but not in memory cells. Overexpression of Blimp-1 in BCL1-3B3 lymphoma cells causes the appearance of an early plasma-cell phenotype, including J-chain expression and IgM secretion^{18–20}. One aspect of the mechanism by which Blimp-1 promotes generation of plasma cells is the repression of c-Myc, thereby allowing the B cell to exit the cell cycle and undergo terminal differentiation^{21–23}.

XBP-1 (X-box-binding protein-1) is a basic-region leucine zipper protein in the CREB/ATF (cyclic AMP response element binding protein/activating transcription factor) family of transcription factors, found ubiquitously in adult tissues but preferentially expressed in fetal exocrine glands, osteoblasts, chondroblasts and liver^{24–26}. XBP-1 is essential for the growth of hepatocytes, as XBP-1-deficient embryos die *in utero* from severe liver hypoplasia and a resulting fatal anaemia²⁷. In human multiple myeloma cells, XBP-1 was selectively induced by treatment with IL-6 and was implicated in the proliferation of malignant plasma cells²⁸. Transient expression studies have shown that the XBP-1 promoter is strongly down-regulated by the transcription factor BSAP/Pax-5, which may account for the high level of XBP-1 expression seen in some plasma cell lines, a developmental stage where BSAP is no longer present²⁹.

As XBP-1-deficient mouse embryos die *in utero*, it was not possible to study immune functions in these animals. Therefore, the RAG-2 (recombination-activating gene-2) complementation system was used to analyse XBP-1-deficient lymphocytes from adult chimaeric animals³⁰. These lymphoid chimaeras displayed a severe defect in the generation of plasma cells and therefore of immunoglobulin secretion. This identifies XBP-1 as a transcription factor essential for the terminal differentiation of B lymphocytes.

High XBP-1 in plasma cells at inflamed sites

We have previously described the regulated expression of XBP-1 transcripts in B cells at various stages of development, and noted the

[§] Present address: Department of Internal Medicine, University of Texas Southwestern Medical Center at Dallas, Texas 75390, USA.

extremely high level of XBP-1 transcripts in several myeloma cell lines²⁹. The disease rheumatoid arthritis offered a source of clinical material to study XBP-1 expression *in vivo* in primary cells in a condition characterized by the proliferation of inflammatory tissue and the infiltration of inflammatory cells, including, in many cases, large numbers of plasma cells³¹. With an *in situ* antisense probe for XBP-1, we demonstrated strong hybridization to plasma cells in rheumatoid synovium from two different patients (Fig. 1a, d), whereas sense probes showed no specific hybridization (Fig. 1b, e). High-power microscopy confirmed the presence of numerous plasma cells in areas of XBP-1 expression (Fig. 1c, f). These data demonstrate high-level expression of XBP-1 transcripts in the plasma cell infiltrates of an inflammatory disease.

Upstream and downstream roles of XBP-1

Activated B cells are driven to become plasma cells by signalling through the CD40 receptor or through mitogens such as lipopolysaccharide (LPS). These two stimuli also drive the upregulation of transcripts encoding XBP-1 in purified B cells (Fig. 2a). The BCL1-3B3 cell line is a highly activated B-cell line that can be driven to an early plasma-cell stage by treatment with IL-2 and IL-5 (refs 32, 33). We tested whether XBP-1 could also drive B-cell differentiation in this model. BCL1-3B3 cells express XBP-1 transcripts at baseline levels, and this does not increase after treatment with IL-2 and IL-5. By analysis with fluorescence-activated cell sorting (FACS), cells overexpressing a bicistronic XBP-1–GFP (green fluorescent protein) construct but not the GFP vector alone showed signs of further maturation: a decrease in CD44 levels and the emergence of

Syndecan-1-positive cells (Fig. 2b). An identical effect has been described for the ectopic expression of transcription factor Blimp-1 in BCL1-3B3 cells¹⁹.

XBP-1/RAG-2-deficient chimaeric mice

XBP-1^{-/-} embryonic stem cells of mice were identified on Southern blots of genomic DNA by the absence of a wild-type XBP-1 band and the presence of two disrupted alleles (Fig. 3a). After injection into RAG-2-deficient blastocysts, XBP-1-deficient cells contributed to reconstitution of peripheral blood B and T lymphocytes in approximately half of the mice born. In these reconstituted animals, XBP-1^{-/-} embryonic stem cells contributed heavily to lymphoid organs, variably to heart, lung, kidney, muscle and bone marrow but minimally to liver (Fig. 3b), consistent with the essential role of XBP-1 in hepatocyte development seen in XBP-1^{-/-} embryos. The reconstitution of lymph nodes and spleen in XBP-1/RAG-2^{-/-} animals resulted in normal-sized organs indistinguishable from those of control 129/SvImJ mice. Chimaeric mice had normal numbers (Fig. 3c), percentages and phenotype of B and T (not shown) lymphocytes on FACS analyses as assessed by expression of B220, IgM and IgD (Fig. 3d). Peritoneal CD5⁺ B-1 lymphocytes were also found in equal numbers in chimaeric and control animals (not shown).

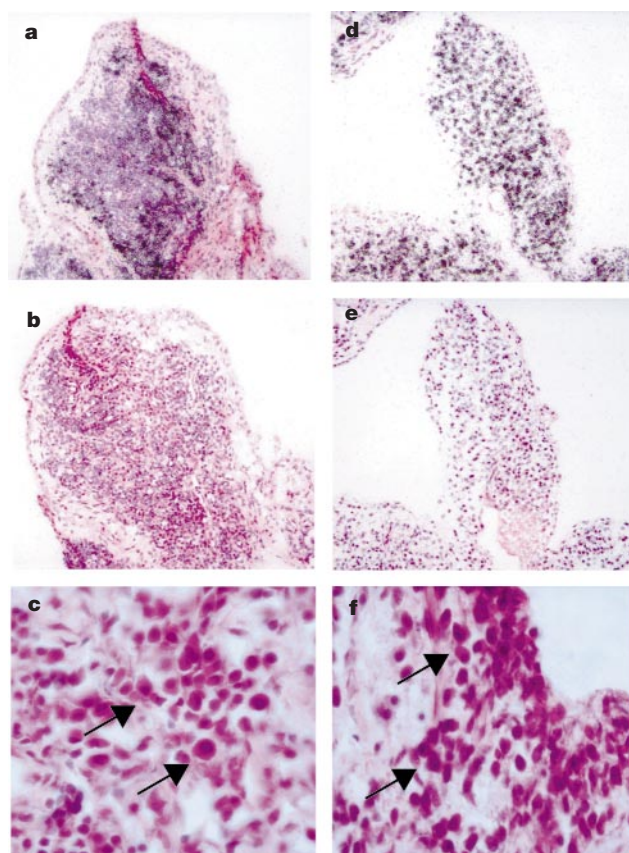


Figure 1 XBP-1 expression in rheumatoid synovium. **a, d**, *In situ* hybridization with an antisense XBP-1 probe on synovium from two patients with rheumatoid arthritis. **b, e**, Sections from the tissue blocks were hybridized with a sense XBP-1 probe (**e**) and with an unrelated (**b**) sense control probe as controls. **c, f**, High-power views of sections stained with haematoxylin and eosin show areas of numerous plasma cells (arrows) in regions where XBP-1 signals were most intense.

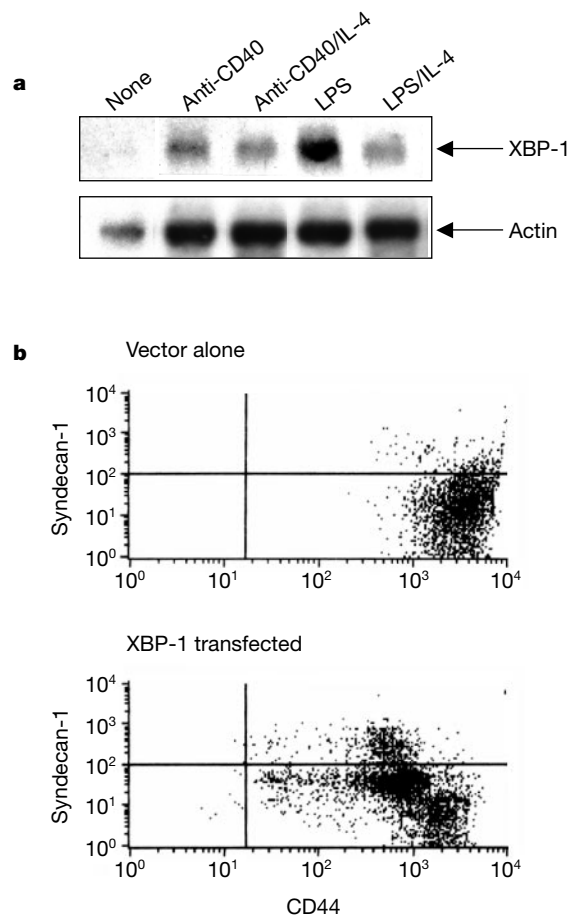


Figure 2 Induction of XBP-1 is upstream and downstream of signals that drive plasma-cell differentiation. **a**, Induction of XBP-1 transcripts in wild-type B lymphocytes. A Northern blot of RNA derived from cells stimulated *in vitro* for 4 d is shown. **b**, XBP-1 drives differentiation of an activated, mature B-cell line. BCL1-3B3 cells were transfected with expression constructs containing no insert (vector alone) or XBP-1 cDNA. Cells were sorted for GFP expression and after 2 d of growth, a FACS analysis was performed. Differentiating cells are recognized by upregulation of cell-surface Syndecan-1 and downregulation of CD44.

Low immunoglobulin from XBP-1^{-/-} B cells

To assess the major function of B lymphocytes, immunoglobulin levels were assayed by enzyme-linked immunosorbent assay (ELISA) in the sera of multiple XBP-1/RAG-2^{-/-} mice and control 129/SvImJ mice. Baseline serum immunoglobulin levels were markedly lower in XBP-1/RAG-2^{-/-} animals for all immunoglobulin subtypes tested, including IgM, with a small amount of IgG2a detected in some of the mice (Fig. 4a). To test the intrinsic potential of XBP-1-deficient B cells for immunoglobulin production after stimulation *in vitro*, splenocytes or purified B cells were treated *in vitro* for 4 days with a stimulus that directly drives immunoglobulin secretion from B cells in the absence of other cell types: the B-cell mitogen LPS. When immunoglobulin levels were determined in culture supernatants, all tested immunoglobulin subtypes were again found at substantially lower levels in XBP-1^{-/-} samples than in control samples (Fig. 4b). To directly demonstrate the effect of XBP-1 in driving immunoglobulin secretion in B cells, retroviral transduction was used to re-express XBP-1 in XBP-1-deficient B cells. Compared with the introduction of the retroviral vector alone, the addition of XBP-1 more than doubled the secretion of IgM (Fig. 4c). Therefore, the XBP-1 transcription factor is very important in the elaboration of immunoglobulins by B lymphocytes.

In vitro phenotype of XBP-1^{-/-} B cells

Given the functional defect in XBP-1^{-/-} terminal B-cell differentiation, we investigated the phenotype of B-cell activation *in vitro*. B cells were stimulated *in vitro* with anti-CD40 antibody, or anti-CD40 and IL-4, for up to 4 days. The B-cell populations were activated similarly, as judged by analysis of the cell-surface markers

MHC class II (I-A), B7.2, CD25, CD69 and CD95 (Fig. 5), and CD19, CD21, CD40, CD44, CD62L and CD80 (not shown). Cell proliferation was the same in XBP-1-deficient and control samples, as judged by equivalent cell counts on days 1–4 of *in vitro* culture (Fig. 6a). The activation of T cells in response to anti-CD3 was also normal as judged by proliferation and expression of CD62L and CD40L (not shown). Therefore, no significant differences were found in the activation of XBP-1-deficient B or T lymphocytes. *In vitro* stimulation with LPS was used to induce class-switch recombination in B cells, a process that is regulated separately for membrane and secreted forms. Class-switch recombination was equivalent in XBP-1^{-/-} and control samples, as judged by FACS analysis of cell-surface immunoglobulin and by polymerase chain reaction with reverse transcription (RT-PCR) with primers that selectively recognize germline IgG2b and IgG3 and all IgM membrane and secreted species (Fig. 6b), thus making a direct role for XBP-1 in regulating immunoglobulin transcription unlikely. However, XBP-1^{-/-} B cells were less differentiated as judged by decreased expression of J-chain, which is required for assembly of IgM and IgA in plasma cells, as well as increased expression of c-Myc, which becomes downregulated as B cells exit the cell cycle to differentiate terminally (Fig. 6c). The cytokine profile of XBP-1-deficient B cells cultured *in vitro* and stimulated *in vitro* with anti-CD40 antibody or LPS, with or without the addition of IL-4, showed no significant differences in the production of IL-6 or IL-10 when compared with control cells (not shown). Therefore, XBP-1-deficient B cells are present in normal numbers and can be activated *in vitro* to undergo proliferation, cell-surface activation-marker expression, class switch recombination, and cytokine secretion at normal levels.

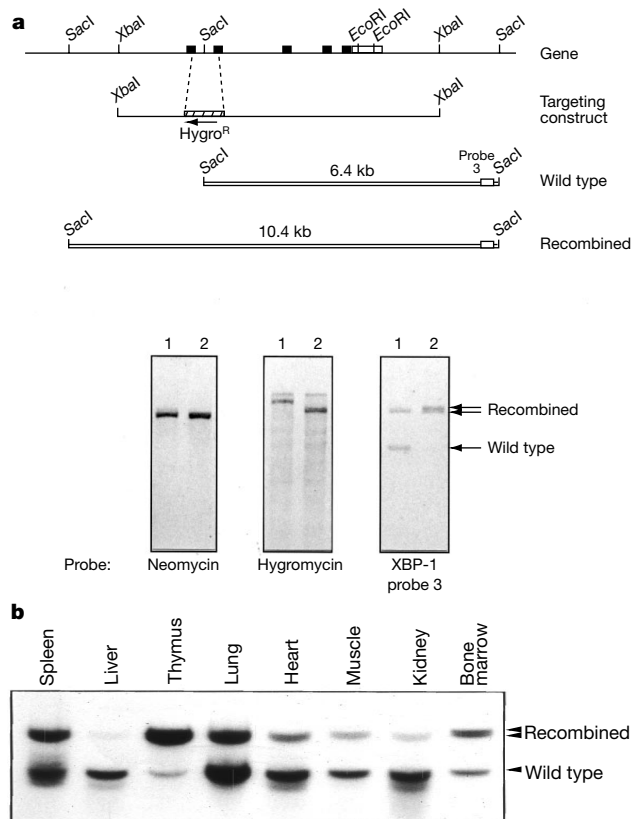
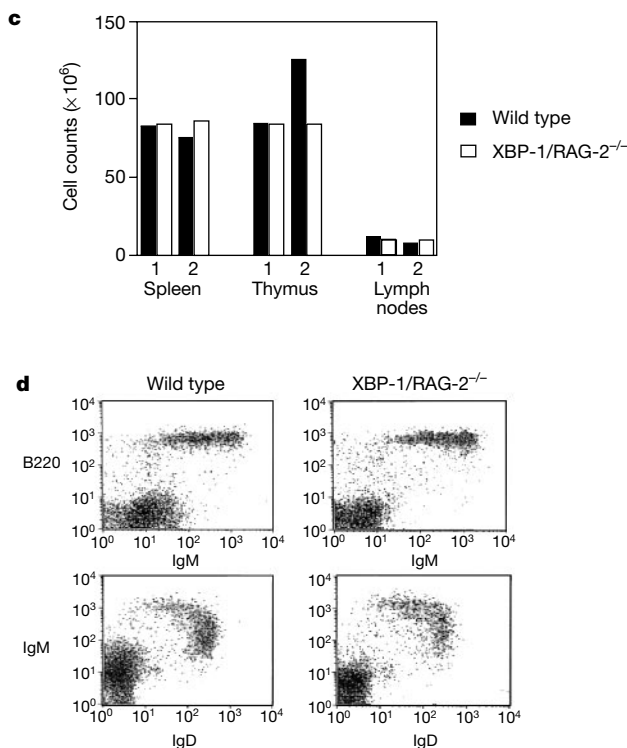


Figure 3 Gene targeting to generate XBP-1-deficient cells. **a**, The XBP-1 gene was disrupted by homologous recombination using a vector containing a hygromycin drug-resistance cassette in place of parts of XBP-1 exons 1 and 2 as well as the intervening intronic DNA. This vector was electroporated into embryonic stem cells already containing a disruption of one XBP-1 allele. **b**, Southern blotting of tissues from mice deficient in XBP



and RAG-2. Cells with a disrupted XBP-1 gene make up much of the lymphoid organs, whereas their contribution to other organs varies widely. **c**, Cell counts of lymphoid organs of XBP-1/RAG-2^{-/-} and wild-type mice. **d**, Phenotype of XBP-1-deficient B lymphocytes. FACS analysis with B220, IgM-FITC and IgD-PE of control 129/SvImJ (wild type) and XBP-1/RAG-2^{-/-} splenocytes.

We conclude that other mechanisms must account for the impaired immunoglobulin production by XBP-1^{-/-} B cells.

Few plasma cells in the absence of XBP-1

The absence of germinal-centre formation can lead to impaired production of immunoglobulin, as evidenced by many mouse mutant models^{34–37}. To assess whether XBP-1-deficient B cells respond to activating stimuli *in vivo*, mice were immunized with 2,4-dinitrophenyl (DNP)-albumin and their draining lymph nodes were gathered after 9 days. Histologic analysis demonstrated normal germinal-centre formation in XBP-1/RAG-2^{-/-} animals (Fig. 7a). This indicates that many of the B-cell functions assayed *in vitro*, such as cell proliferation and activation, were also intact *in vivo*.

Histological analysis of organs from XBP-1/RAG-2^{-/-} animals, however, revealed a striking absence of plasma cells. Sections of jejunum from XBP-1/RAG-2^{-/-} and control mice showed approximately 70-fold fewer plasma cells in the former (Fig. 6b). Twenty high-power fields (×50 objective) of lymphoid tissue in spleen were examined from three immunized wild-type mice and three immunized XBP-1^{-/-} chimaeric mice. The average number of plasma cells in the wild-type mice per field was 5, whereas the average number in the knockout mice was 0.2. In addition, it was noted that the rare plasma cells seen in XBP-1/RAG-2^{-/-} animals generally lacked a perinuclear huff, indicating that these cells did not have a large Golgi apparatus and were probably not secreting large amounts of immunoglobulins. Histological examination of spleen and bone

marrow revealed a similar striking absence of plasma cells in XBP-1/RAG-2^{-/-} animals (not shown).

An additional indicator of plasma-cell differentiation is the presence of cells positive for Syndecan-1. Syndecan-1 is a cell-surface glycoprotein upregulated on terminally differentiated plasma cells but absent on mature B cells³⁸. As predicted by the low level of immunoglobulin secretion, XBP-1^{-/-} samples from mice immunized with DNP-albumin did not contain significant numbers of Syndecan-1-positive cells, whereas a significant increase above baseline was readily detected in the wild-type control (Fig. 7c). Therefore, XBP-1/RAG-2^{-/-} animals display a severe defect in the generation of terminally differentiated plasma cells. The finding of very few plasma cells despite apparently normal germinal centres locates the primary site of XBP-1 action to the interval between full B-cell activation and terminal differentiation. It is not possible to more precisely localize the block as very few markers are available that distinguish activated mature B cells from early plasma cells.

XBP-1^{-/-} B cells are unresponsive to antigen *in vivo*

Whereas the above experiments firmly establish an intrinsic defect in production of antibodies by B cells when activated *in vitro*, it is not clear to what extent this defect is relevant to B-cell function *in vivo*. The potential of XBP-1^{-/-} B cells to secrete both T-cell-independent and -dependent antigen-specific immunoglobulin was therefore assessed by immunization of mice with TNP-Ficoll and TNP-CGG (chicken gamma-globulin), respectively. Antigen-specific immunoglobulin production by XBP-1^{-/-} B cells in response to the T-cell-independent antigen TNP-Ficoll, which was primarily of the IgM and IgG3 isotypes, was measured on days 7, 14 and 21 after immunization and found to be virtually undetectable (Fig. 8a). This defect was equally apparent in mice immunized with the T-cell-dependent antigen TNP-CGG (Fig. 8b) at the same time

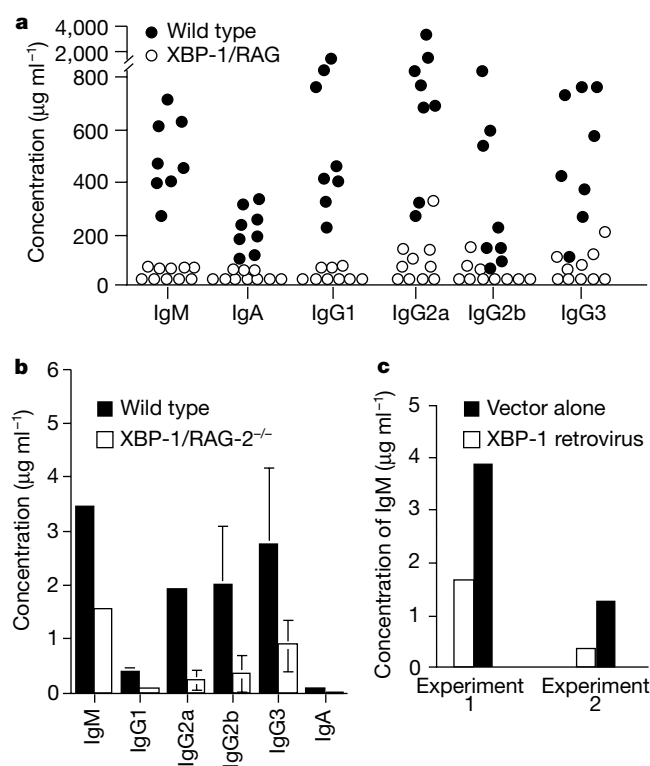


Figure 4 XBP-1 is required for immunoglobulin secretion. **a**, Baseline serum immunoglobulin levels in age-matched control 129/SvImJ mice or XBP-1/RAG-2^{-/-} mice as measured by ELISA. **b**, B lymphocytes from 129/SvImJ or XBP-1/RAG-2^{-/-} animals were cultured *in vitro* with LPS, and immunoglobulin levels in the supernatants were determined after 4 d of culture. **c**, Partial restoration of immunoglobulin secretion by XBP-1^{-/-} B cells. After activation and transduction with a retroviral vector alone or with a retrovirus expressing XBP-1, cells were FACS sorted 24 h after transduction to include only GFP-positive cells that were actively transcribing the retroviral construct. XBP-1-deficient B cells were cultured *in vitro* for 3 days further and the supernatants were tested for IgM concentration.

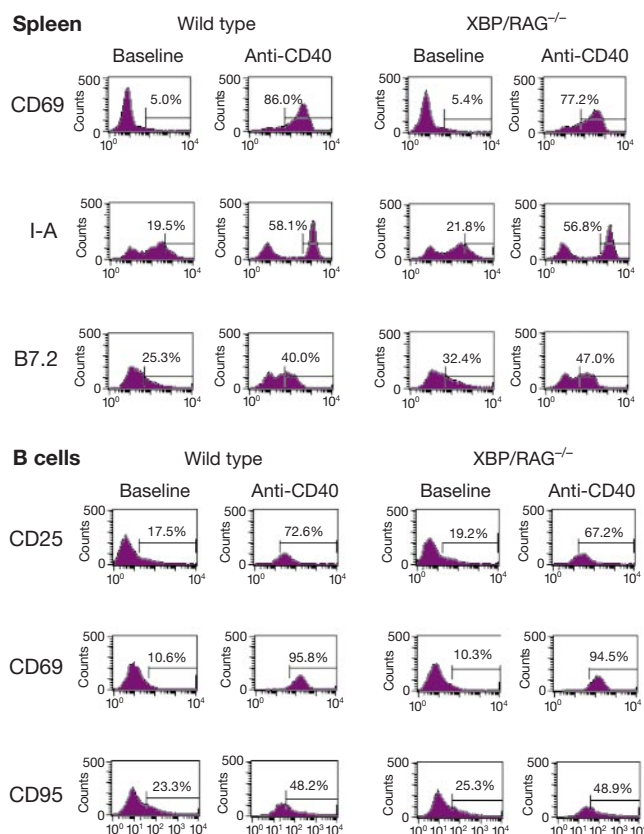


Figure 5 Surface activation markers (CD69, MHC class II (I-A), B7.2, CD25 and CD95) on B cells are normal in the absence of XBP-1.

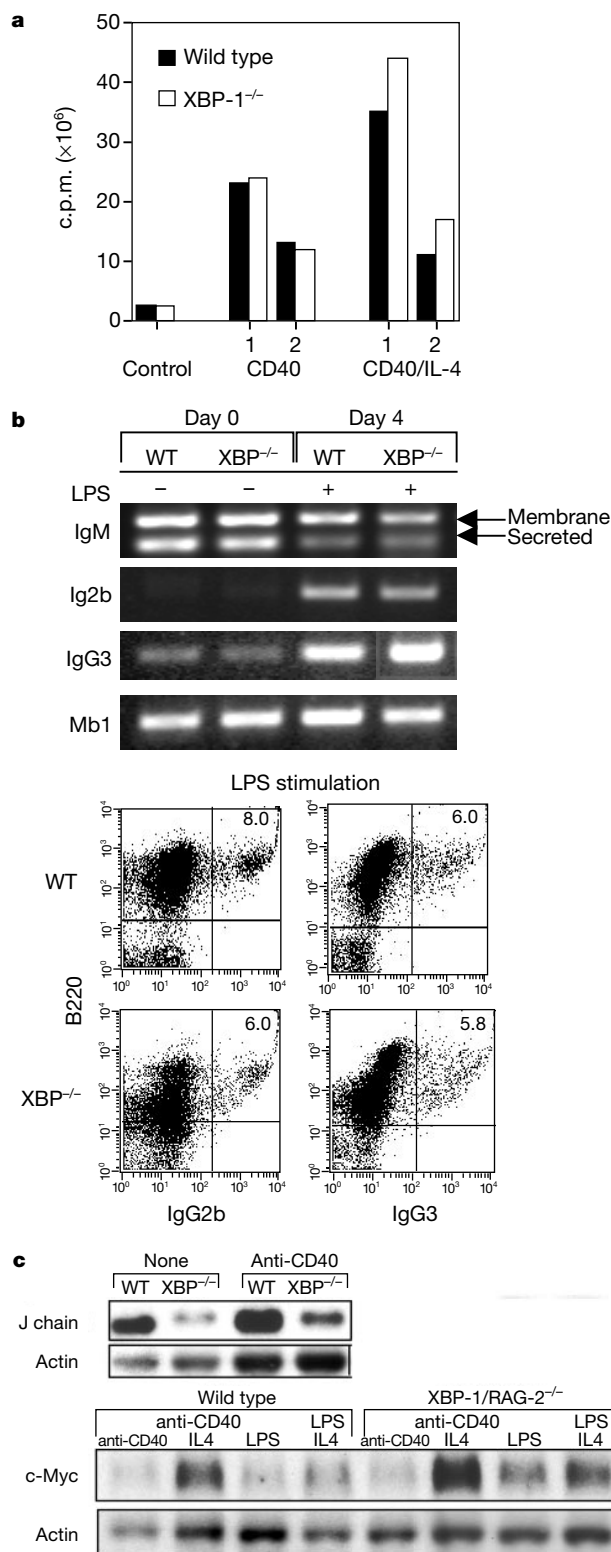


Figure 6 *In vitro* phenotype of XBP-1^{-/-} B cells. **a**, B-cell proliferation is normal in XBP-1^{-/-} cells. **b**, Class switch recombination is unaffected by the absence of XBP-1. FACS analysis of surface IgG, and RT-PCR analysis of all membrane and secreted IgM transcripts and IgG2b and IgG3 germline transcripts after LPS stimulation. WT, wild type. **c**, Diminished terminal differentiation of XBP-1-deficient B lymphocytes after *in vitro* stimulation for 4 d. Northern blots show that control cultures have undergone further differentiation than XBP-1-deficient cultures as XBP-1^{-/-} B cells have decreased levels of J-chain transcripts and increased levels of c-Myc transcripts. B lymphocytes from 129/SvImJ or XBP-1/RAG-2^{-/-} animals were cultured *in vitro* with anti-CD40 antibody, anti-CD40 plus IL-4, LPS or LPS plus IL-4.

points and involved all responding immunoglobulin subclasses including the IgM, IgG1 and IgG2a isotypes. These findings indicate that, although XBP-1-deficient B cells are generated in normal numbers, they fail to become antibody-secreting cells *in vivo* in response to antigen-specific, T-cell-independent or -dependent stimuli. The failure to respond to antigens that are not dependent on T-cell help clearly indicates an intrinsic B-cell defect in the secretion of immunoglobulin in the absence of XBP-1.

XBP-1^{-/-} chimaeras are unresponsive to polyoma virus

We have determined that XBP-1^{-/-} chimaeras do not mount an antibody response when immunized with either T-cell-independent or -dependent antigens, and established by *in vitro* assays that there is an intrinsic defect in the B-cell compartment. We next determined whether an immune response against pathogens that evoke protective humoral immune responses in addition to cell-mediated immune responses could be mounted in the absence of XBP-1. Viruses can be controlled exclusively by antibodies (rotavirus, polyoma virus, vesicular stomatitis virus), by antibodies and T cells (influenza, poliovirus, encephalomyocarditis), or exclusively by T cells (lymphocyte choriomeningitis virus, human immuno-

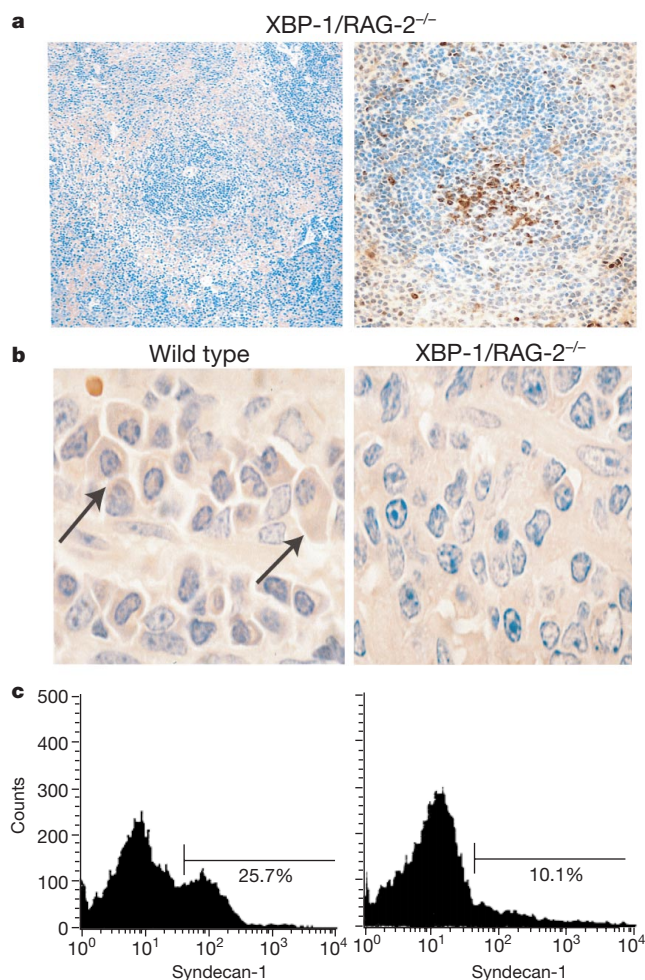


Figure 7 Decreased formation of plasma cells in the absence of XBP-1. **a**, Histological section of a germinal centre from a draining lymph node of an XBP-1/RAG-2^{-/-} mouse immunized with DNP-albumin. The histologic appearance is normal. **b**, Sections of mouse jejunum showing a marked reduction in plasma cell numbers in XBP-1/RAG-2^{-/-} animals. **c**, Lack of plasma-cell generation after immunization of XBP-1/RAG-2^{-/-} mice. FACS analysis of B lymphocytes from control (left) and XBP-1/RAG-2^{-/-} (right) animals showed an increase in Syndecan-1-positive cells (a marker for plasma cells) only in control samples.

deficiency virus, simian immunodeficiency virus)³⁹. We therefore determined the IgM and IgG response of XBP-1^{-/-} chimaeras to polyoma virus, a virus that is controlled by B cells, apparently in the absence of T-cell help⁴⁰. B cells purified from wild-type mice or from XBP-1/RAG^{-/-} chimaeras (or from irradiated recipients reconstituted with bone marrow or spleen from such chimaeras) were transferred into host severe combined immunodeficient (SCID) mice that were then infected with polyoma virus. As shown, XBP-1^{-/-}-transferred B cells did not confer protection to polyoma virus as evidenced by the failure of these reconstituted recipients to produce polyoma-virus-specific (VP1-specific) IgG (Fig. 8c) or IgM

(not shown). These mice also had greatly impaired survival when compared with the control recipients.

Discussion

XBP-1 is the only transcription factor known to be selectively and specifically required for the terminal differentiation of B lymphocytes to plasma cells. Its induction on B-cell activation and high-level expression in plasma cells coupled with the absence of plasma cells in XBP-1^{-/-} lymphoid chimaeras identifies a single important site of action for XBP-1 in the terminal stages of the B lineage. Significantly, germinal centres were formed in the absence of XBP-1, placing the observed defect distal to the actions of several transcription factors shown to be essential for germinal-centre formation, including Bcl-3, Bcl-6, NF- κ B/p52 and IRF-4 (refs 34–37). B cells that survive the selection, activation and differentiation steps that occur in germinal centres exit to become either memory B cells or plasma cells. XBP-1 is now known to be essential for this process, although other molecular details remain unknown. Recent histological studies have indicated that some cells destined to become plasma cells may already be identifiable in the germinal centre. The transcription factors Blimp-1 and IRF-4 are heavily expressed in plasma cells but have also been detected by specific antibodies in a small subset of light-zone germinal-centre B cells, probably representing cells exiting the germinal centre to differentiate to plasma cells^{20,35}. Morphologically, such cells have had the phenotype of centrocytes or the more mature plasmablasts or plasma cells. Neither Blimp-1 nor IRF-4 is expressed in most germinal-centre B cells, indicating specific upregulation during terminal differentiation. However, IRF-4 is already known to be essential for the earlier steps of germinal-centre formation, and the role of Blimp-1 *in vivo* is still being elucidated.

We found that XBP-1^{-/-} B cells could be activated but were blocked in terminal differentiation, a phenotype characterized by ongoing transcription of the proto-oncogene *c-myc*. XBP-1 and Blimp-1 have identical effects when overexpressed *in vitro*, leading to differentiation of Bcl-1 lymphoma cells but also to apoptosis of immature B-cell lines (refs 41 and 42; A.M.R. and L.H.G., unpublished observations). Such differences in responses, dependent on the stage of B-cell differentiation, have previously been noted following B-cell-receptor ligation. The underlying differences in signal transduction between immature and mature B cells remain poorly understood: B-cell-receptor ligation results in cell-cycle arrest and apoptosis of stimulated immature B cells^{7,22,42}, whereas mature B cells initiate expression of *Egr-1*, *c-Fos* and *c-Myc* and go on to activate and proliferate. Subsequent downregulation of the *c-myc* gene allows B cells to cease proliferating and to differentiate, instead. The repression of the *c-myc* promoter is an important mechanism of action for the complex of Blimp-1 and histone deacetylase in contributing to B-cell differentiation^{21,23}. However, exiting the cell cycle is not sufficient to cause B-cell differentiation⁴³. This indicates that as-yet-unknown XBP-1 or Blimp-1 target genes are essential in plasma-cell generation. However, the functions of XBP-1 and Blimp-1 are unique, as levels of Blimp-1 messenger RNA in activated B cells were not affected by the absence of XBP-1 (not shown) and thus Blimp-1 does not compensate for the absence of XBP-1. Further, XBP-1 does not directly repress activation of the *c-myc* promoter (A.M.R. and L.H.G., unpublished observations). Finally, whether Blimp-1 is involved in T-cell-dependent as well as T-cell-independent antibody production is controversial, whereas XBP-1 clearly is vital for both of these processes^{21,44}. Therefore, XBP-1 acts downstream of Blimp-1 or through a separate pathway in its regulation of B-cell differentiation.

Our experiments therefore indicate that XBP-1 regulates a final common pathway of plasma-cell generation, regardless of the immune stimulus or the anatomic location of the immune response. Immunoglobulin production remained minimal in the absence of XBP-1, whether animals were immunized with T-cell-independent

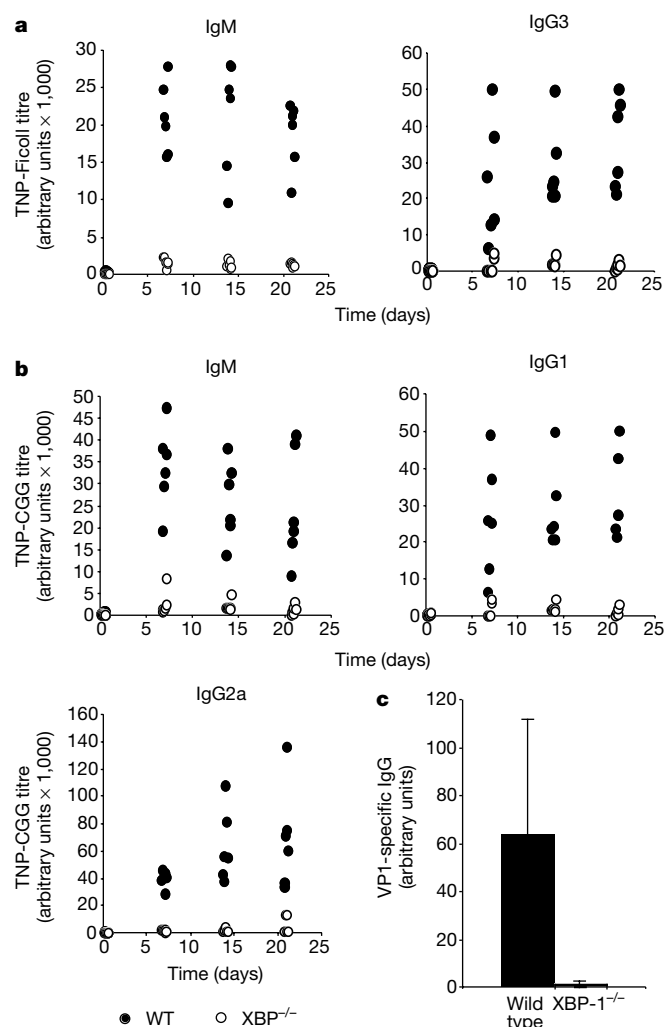


Figure 8 XBP-1^{-/-} lymphocytes do not respond to T-cell-independent or -dependent antigens or to polyoma virus *in vivo*. **a**, T-cell-independent antibody production. Serum antibodies in response to immunizations with TNP-Ficoll were determined in wild-type 129/Sv and XBP-1/RAG-2^{-/-} mice. Mice were immunized with 50 μ g of TNP-Ficoll precipitated in alum. Serum was analysed for TNP-specific IgM and IgG3 antibodies on days 7, 14 and 21. Data points are arbitrary units of optical density for individuals ($n = 6$). **b**, T-cell-dependent antibody production. Serum antibodies in response to immunizations with TNP-CGG were determined in wild-type 129/Sv and XBP-1/RAG-2^{-/-} mice. Mice were immunized with 100 μ g of TNP-CGG precipitated in alum. Serum was analysed for TNP-specific IgM, IgG1 and IgG2a antibodies on days 7, 14 and 21. Data points are arbitrary units of optical density for individuals ($n = 6$). **c**, Polyoma-virus-specific IgG response of SCID mice reconstituted with XBP-1^{-/-} B cells. VP1-specific IgG (day 21; tested in 1:50 to 1:400 dilution) of SCID mice reconstituted with T-cell-depleted splenocytes of control 129/Sv ($n = 3$) or XBP-1/RAG-2^{-/-} ($n = 3$) mice. Values are expressed as arbitrary optical-density units at 450 nm, with background values (obtained from uninfected mouse serum) subtracted. The optical-density values obtained with uninfected mouse serum are below 0.08.

antigen, T-cell-dependent antigen or polyoma virus. Furthermore, the populations of plasma cells that are normally found in lymph nodes, spleen, bone marrow or lamina propria of the gut were all severely reduced in the absence of XBP-1, indicating that the action of XBP-1 is fundamental to the mechanism of plasma-cell generation. We conclude that XBP-1 represents a molecular switch in the terminal differentiation of B cells. XBP-1 clearly is intrinsic in B-cell differentiation, yet low-level expression of XBP-1 is ubiquitous so there may well be additional roles for XBP-1 in the immune system, a possibility that we are currently exploring. □

Methods

Disruption of XBP-1

A genomic clone containing XBP-1 was disrupted by deletion of parts of exons 1 and 2 along with the intervening intron and by insertion of a cassette encoding a gene for hygromycin resistance. Embryonic stem cells from the 129 mouse strain with one allele of XBP-1 previously disrupted by the neomycin gene were transfected with the new construct and selected in hygromycin, neomycin and gancyclovir²⁷. Surviving clones were screened by Southern blotting for the presence of two disrupted XBP-1 alleles and the loss of the wild-type allele. Appropriate embryonic-stem clones were injected into RAG-2-deficient blastocysts and implanted into pseudopregnant female mice as described³⁰. The resulting mice were assayed for lymphoid reconstitution by FACS analysis of peripheral blood mononuclear cells using anti-CD3 and anti-B220 antibodies (PharMingen).

Antibodies and cell culture

FACS analysis was performed using antibodies to mouse CD69, MHC class II (I-A), CD3, B220, CD44 and Syndecan-1 (PharMingen). The BCL1-3B3 cell line (ATCC) was grown according to the supplied protocol. Transfection of the BCL1-3B3 cell lines was by electroporation at 290 V and 275 μ F in a BioRad Electroporator. Cells were then sorted for GFP expression and after 2 d of growth, a FACS analysis was performed. Mouse splenocytes or purified B cells (B220⁺ magnetic bead selection, Miltenyi Biotech) were plated at 10^6 cells ml^{-1} and stimulated for up to 4 d with anti-CD40 (1 $\mu\text{g ml}^{-1}$), anti-CD40 plus IL-4 (10 ng ml^{-1}), LPS (20 $\mu\text{g ml}^{-1}$), or LPS plus IL-4.

ELISA assays

Assays of immunoglobulin or cytokine levels in serum or in culture supernatants were performed as described⁴⁵.

In situ hybridization

Samples of synovial tissue from patients with the clinical diagnosis of rheumatoid arthritis were collected at the time of joint arthroplasty as discarded materials. Tissue collection was approved by the Institutional Review Board. *In situ* hybridization was performed with an XBP-1-specific riboprobe as described^{15,26}.

Retroviral transduction of B cells

The XBP-1 complementary DNA was cloned into the retroviral vector pRV-GFP, which contains the gene for green fluorescent protein expression. A calcium-phosphate transfection was used to introduce the DNA into the Phoenix cell line as described⁴⁶. Viral supernatants were collected after 48 h and frozen at -80°C for later use. B cells were purified from mouse spleens using positive selection on B220⁺ magnetic beads as described by the manufacturer (MidiMacs, Miltenyi Biotech). B-cell purity was generally near 95%. The B cells were then activated in culture with LPS (25 $\mu\text{g ml}^{-1}$) for 24 h. The cells were spun out, mixed with 4 mg of polybrene and 1 ml of virus-containing supernatant and spun at 6,400 r.p.m. for 30 min at 32°C . Cells were sorted for GFP expression and were grown for a further 1–4 d in culture, with or without further activating stimuli.

PCR amplification of immunoglobulin transcripts

Total RNA was isolated on day 3 or 4 from 10^7 LPS-stimulated splenocytes using the Trizol reagent (GIBCO/BRL) as per the manufacturer's instructions. We generated cDNA by using Superscript (GIBCO/BRL) according to the manufacturer's instructions. For IgM the primers used were 5'-TCCTCTGAGCCTCTTCTAC-3' and 5'-CCAGACATTCCTTCAGATTG-3' (membrane) and 5'-CACACTGTACAATGTCTCCCT-3' and 5'-AAAATGCAACATCTCATCTCTG-3' (secreted transcripts), sense and antisense, respectively. Germline IgG2b, IgG3 and mb1 transcripts were amplified with methods and primers as described⁴⁷.

Flow cytometry analysis

Single-cell suspensions from spleens were prepared as described⁴⁷. Briefly, cells from day 4 or 5 cultures were washed in PBS with 2% FCS and stained with various antibodies conjugated with phycoerythrin (IgG2b, IgG3) and Cychrome (B220) (PharMingen). The cells were analysed with FACScalibur (Becton Dickinson, Sparks) and Cellquest software, and are presented as dot plots or histograms after gating for live cells.

Immunization

Wild-type 129/Sv mice and XBP-1/RAG^{-/-} chimaeric mice (10–13 weeks old) were

immunized as described⁴⁸ with a single dose of 100 μg of TNP-CGG or 50 μg TNP-Ficoll (Biosearch) that had been precipitated in alum. All antigens were resuspended in saline and injected intraperitoneally.

Polyoma virus experiments

Adoptive transfer of B-cell-containing splenocyte populations of wild-type and mutant (XBP-1^{-/-}) mice into C57BL/SCID mice was performed as described⁴⁹. Briefly, spleen cell suspensions were obtained by homogenizing spleens between frosted-glass microscope slides, and the erythrocytes were lysed by treatment with 0.83% ammonium chloride. After *in vitro* T-cell depletion with rat anti-mouse Thy1.2 antibody (PharMingen) and rabbit complement (Pel-Freez, Brown Deer), the number of viable cells was counted and the purity (99%) assessed by FACS. Spleens of 129/Sv and XBP-1^{-/-} mice yielded comparable numbers of cells after T-cell depletion and in the 129/Sv and XBP-1^{-/-} adoptive transfer experiments, 5×10^7 cells were injected intravenously into each SCID mouse.

Virus Ag-specific ELISA assays were done using purified VP1 PyV capsid Ag (50 ng ml^{-1}) that was produced by recombinant baculovirus expression vectors in Sf9 insect cells and purified (a gift of R. Garcea). The serum samples were tested in duplicate using biotinylated goat anti-mouse IgM or IgG and streptavidin-HRP (Vector, Burlingame) to detect IgM or IgG, respectively.

Received 29 March; accepted 24 May 2001.

- Ahmed, R. & Gray, D. Immunological memory and protective immunity: understanding their relation. *Science* **272**, 54–60 (1996).
- DalPorto, J. M., Haberman, A. M., Shlomchik, M. J. & Kelsoe, G. Antigen drives very low affinity B cells to become plasmacytes and enter germinal centers. *J. Immunol.* **161**, 5373–5381 (1998).
- Arpin, C. *et al.* Generation of memory B cells and plasma cells *in vitro*. *Science* **268**, 720–722 (1995).
- Foy, T. M., Aruffo, A., Bajorath, J., Buhlmann, J. E. & Noelle, R. J. Immune regulation by CD40 and its ligand GP39. *Annu. Rev. Immunol.* **14**, 591–617 (1996).
- Randall, T. D. *et al.* Interleukin-5 (IL-5) and IL-6 define two molecularly distinct pathways of B-cell differentiation. *Mol. Cell. Biol.* **13**, 3929–3936 (1993).
- Takatsu, K. Cytokines involved in B-cell differentiation and their sites of action. *Proc. Soc. Exp. Biol. Med.* **215**, 121–133 (1997).
- Altmeyer, A. *et al.* Reversal of EBV immortalization precedes apoptosis in IL-6-induced human B cell terminal differentiation. *Immunity* **7**, 667–677 (1997).
- Illera, V. A., Perandones, C. E., Stunz, L. L., Mower, D. A. Jr & Ashman, R. F. Apoptosis in splenic B lymphocytes. Regulation by protein kinase C and IL-4. *J. Immunol.* **151**, 2965–2973 (1993).
- Choe, J. & Choi, Y. S. IL-10 interrupts memory B cell expansion in the germinal center by inducing differentiation into plasma cells. *Eur. J. Immunol.* **28**, 508–515 (1998).
- Liu, Y. J. & Banachereau, J. Regulation of B-cell commitment to plasma cells or to memory B cells. *Semin. Immunol.* **9**, 235–240 (1997).
- McHeyzer-Williams, M. G. & Ahmed, R. B cell memory and the long-lived plasma cell. *Curr. Opin. Immunol.* **11**, 172–179 (1999).
- Morse, L., Chen, D., Franklin, D., Xiong, Y. & Chen-Kiang, S. Induction of cell cycle arrest and B cell terminal differentiation by CDK inhibitor p18^{INK4} and IL-6. *Immunity* **6**, 47–56 (1997).
- Kopf, M. *et al.* Impaired immune and acute-phase responses in interleukin-6-deficient mice. *Nature* **368**, 339–342 (1994).
- Henderson, A. & Calame, K. Transcriptional regulation during B cell development. *Annu. Rev. Immunol.* **16**, 163–200 (1998).
- Glimcher, L. H. & Singh, H. Transcription factors in lymphocyte development—T and B cells get together. *Cell* **96**, 13–23 (1999).
- Oliver, A. M., Martin, F. & Kearney, J. F. Mouse CD38 is down-regulated on germinal center B cells and mature plasma cells. *J. Immunol.* **158**, 1108–1115 (1997).
- Morrison, A. M., Nutt, S. L., Thevenin, C., Tolink, A. & Busslinger, M. Loss- and gain-of-function mutations reveal an important role of BSAP (Pax-5) at the start and end of B cell differentiation. *Semin. Immunol.* **10**, 133–142 (1998).
- Schliephake, D. E. & Schimpl, A. Blimp-1 overcomes the block in IgM secretion in lipopolysaccharide/anti-mu (Fab')₂-co-stimulated B lymphocytes. *Eur. J. Immunol.* **26**, 268–271 (1996).
- Turner, C. A. Jr, Mack, D. H. & Davis, M. M. Blimp-1, a novel zinc finger-containing protein that can drive the maturation of B lymphocytes into immunoglobulin-secreting cells. *Cell* **77**, 297–306 (1994).
- Angelini-Duclos, C., Cattoretti, G., Lin, K.-I. & Calame, K. Commitment of B lymphocytes to a plasma cell fate is associated with Blimp-1 expression *in vivo*. *J. Immunol.* **165**, 5462–5471 (2000).
- Lin, Y., Wong, K.-K. & Calame, K. Repression of c-myc transcription by Blimp-1, an inducer of terminal B cell differentiation. *Science* **276**, 596–599 (1997).
- Melchers, F. B lymphocyte-lineage cells from early precursors to Ig-secreting plasma cells: targets of regulation by the myc/mad/max families of genes? *Curr. Top. Microbiol. Immunol.* **224**, 19–30 (1997).
- Yu, J., Angelini-Duclos, C., Greenwood, J., Liao, J. & Calame, K. Transcriptional repression by Blimp-1 (PRDI-BF1) involves recruitment of histone deacetylase. *Mol. Cell. Biol.* **20**, 2592–2603 (2000).
- Liou, H. C. *et al.* A new member of the leucine zipper class of proteins that binds to the HLA DR alpha promoter. *Science* **247**, 1581–1584 (1990).
- Clauss, I. M. *et al.* In situ hybridization studies suggest a role for the basic region-leucine zipper protein hXBP-1 in exocrine gland and skeletal development during mouse embryogenesis. *Dev. Dynam.* **197**, 146–156 (1993).
- Gravallese, E. M., Darling, J. M., Ladd, A. L., Katz, J. N. & Glimcher, L. H. *In situ* hybridization studies of stromelysin and collagenase mRNA expression in rheumatoid synovium. *Arthritis Rheum.* **34**, 1076–1084 (1991).
- Reimold, A. M. *et al.* An essential role in liver development for transcription factor XBP-1. *Genes Dev.* **14**, 152–157 (2000).
- Wen, X. Y. *et al.* Identification of c-myc promoter-binding protein and X-box binding protein 1 as interleukin-6 target genes in human multiple myeloma cells. *Int. J. Oncol.* **15**, 173–178 (1999).
- Reimold, A. M. *et al.* Transcription factor B cell lineage-specific activator protein regulates the gene for human X-box binding protein 1. *J. Exp. Med.* **183**, 393–401 (1996).
- Chen, J., Lansford, R., Stewart, V., Young, F. & Alt, F. W. RAG-2-deficient blastocyst complementation: an assay of gene function in lymphocyte development. *Proc. Natl Acad. Sci. USA* **90**, 4528–4532 (1993).

31. Kim, H. J., Krenn, V., Steinhauser, G. & Berek, C. Plasma cell development in synovial germinal centers in patients with rheumatoid and reactive arthritis. *J. Immunol.* **162**, 3053–3062 (1999).
32. Blackman, M. A., Tigges, M. A., Minie, M. E. & Koshland, M. E. A model system for peptide hormone action in differentiation: interleukin 2 induces a B lymphoma to transcribe the J chain gene. *Cell* **47**, 609–617 (1986).
33. Brooks, K., Yuan, D., Uhr, J. W., Krammer, P. H. & Vitetta, E. S. Lymphokine-induced IgM secretion by clones of neoplastic B cells. *Nature* **302**, 825–826 (1983).
34. Franzoso, G. *et al.* Mice deficient in nuclear factor (NF)-kappa B/p52 present with defects in humoral responses, germinal center reactions, and splenic microarchitecture. *J. Exp. Med.* **187**, 147–159 (1998).
35. Mittrucker, H. W. *et al.* Requirement for the transcription factor LSIRF/IRF4 for mature B and T lymphocyte function. *Science* **275**, 540–543 (1997).
36. Schwarz, E. M., Krimpenfort, P., Berns, A. & Verma, I. M. Immunological defects in mice with a targeted disruption in Bcl-3. *Genes Dev.* **11**, 187–197 (1997).
37. Dent, A. L., Shaffer, A. L., Yu, X., Allman, D. & Staudt, L. M. Control of inflammation, cytokine expression, and germinal center formation by BCL-6. *Science* **276**, 589–592 (1997).
38. Sanderson, R. D., Lalor, P. & Bernfield, M. B lymphocytes express and lose syndecan at specific stages of differentiation. *Cell Regul.* **1**, 27–35 (1989).
39. Bachmann, M. F. & Kopf, M. The role of B cells in acute and chronic infections. *Curr. Opin. Immunol.* **11**, 332–339 (1999).
40. Szomolanyi-Tsuda, E. & Welsh, R. M. T cell-independent antibody-mediated clearance of polyoma virus in T cell-deficient mice. *J. Exp. Med.* **183**, 403–411 (1996).
41. Messika, E. J. *et al.* Differentiated effect of B lymphocyte-induced maturation protein (Blimp-1) expression on cell fate during B cell development. *J. Exp. Med.* **188**, 515–525 (1998).
42. Knodel, M., Kuss, A. W., Lindemann, D., Berberich, I. & Schimpl, A. Reversal of Blimp-1-mediated apoptosis by A1, a member of the Bcl-2 family. *Eur. J. Immunol.* **29**, 2988–2998 (1999).
43. Lin, K.-I., Lin, Y. & Calame, K. Repression of c-myc is necessary but not sufficient for terminal differentiation of B lymphocytes in vitro. *Mol. Cell. Biol.* **20**, 8684–8695 (2000).
44. Soro, P. G. *et al.* Differential involvement of the transcription factor blimp-1 in T cell-independent and -dependent B cell differentiation to plasma cells. *J. Immunol.* **163**, 611–617 (1999).
45. Hodge, M. R. *et al.* Hyperproliferation and dysregulation of IL-4 expression in NF-ATp-deficient mice. *Immunity* **4**, 397–405 (1996).
46. Szabo, S. J. *et al.* A novel transcription factor, T-bet, directs Th1 lineage commitment. *Cell* **100**, 655–669 (2000).
47. Manis, J. P. *et al.* Class switching in B cells lacking 3' immunoglobulin heavy chain enhancers. *J. Exp. Med.* **188**, 1421–1431 (1998).
48. Gartner, E., Alt, F. W., Monroe, R. J. & Seidl, K. J. Antigen-independent appearance of recombination activating gene (RAG)-positive bone marrow B cells in the spleens of immunized mice. *J. Exp. Med.* **192**, 1745–1754 (2000).
49. Szomolanyi-Tsuda, E., Brien, J. D., Dorgan, J. E., Welsh, R. M. & Garcea, R. L. The role of CD40–CD154 interaction in antiviral T cell-independent IgG responses. *J. Immunol.* **164**, 5877–5882.

Acknowledgements

This work was supported by grants from the National Institutes of Health (L.H.G., F.A. and A.R.) and the Hood Foundation (J.M.), and a gift from the G. Harold and Leila Y. Mathers Charitable Foundation (L.H.G.). We thank A. Erlebacher, K. Mowen and S. Peng for a review of the manuscript; L. Davidson and K. Sigrist for production of XBP/RAG^{-/-} chimaeras; M. Wheaton for technical assistance with the polyoma virus experiments; C. McCall for preparation of the manuscript; and A. Bottaro and I. Kuzin for IgM primers.

Correspondence and requests for materials should be addressed to L.H.G. (e-mail: lglimche@hsph.harvard.edu).

A massive cloud of cold atomic hydrogen in the outer Galaxy

Lewis B. G. Knee* & Christopher M. Brunt*†

* National Research Council of Canada, Herzberg Institute of Astrophysics, Dominion Radio Astrophysical Observatory, PO Box 248, Penticton, British Columbia, V2A 6K3, Canada

† Department of Physics and Astronomy, University of Calgary, 2500 University Drive NW, Calgary, Alberta, T2N 1N4, Canada

A large fraction of the mass of the interstellar medium in our Galaxy is in the form of warm (10^3 – 10^4 K) and cool (50–100 K) atomic hydrogen (H I) gas¹. Cold (10–30 K) regions are thought to be dominated by dense clouds of molecular hydrogen². Cold H I is difficult to observe, and therefore our knowledge of its abundance and distribution in the interstellar medium is poor. The few known clouds of cold H I are much smaller in size and mass than typical molecular clouds^{3–5}. Here we report the discovery that the H I supershell GSH139–03–69 is very cold (10 K). It is about 2 kiloparsecs in size and as massive as the largest molecular complexes⁶. The existence of such an immense structure composed of cold atomic hydrogen in the interstellar medium runs counter to the prevailing view that cold gas resides almost exclusively in clouds dominated by molecular hydrogen.

Clouds of cold H I are usually detected through their absorption of the H I line radiation emitted by brighter background H I gas⁷ ('self-absorption'). The angular size of most clouds that have been identified by self-absorption is limited by the selection bias that the bright H I background be relatively uniform. This has precluded the detection of physically large structures of cold H I that might exist in the Galaxy. A further limiting factor has been the low angular resolution ($\sim 0.5^\circ$) of existing large-scale Galactic H I surveys. The other method of finding cold H I, line absorption against background radio-continuum sources⁷, can only set upper limits on temperatures in the cold gas⁸ and there are usually too few continuum sources behind a given cloud to permit mapping.

Our H I line (frequency 1.420 GHz) data were obtained during the

Canadian Galactic Plane Survey (CGPS) using the Synthesis Telescope⁹ at the Dominion Radio Astrophysical Observatory. Details of the observations and data processing are given elsewhere^{10,11}. In Fig. 1 we display our data as a mosaic. The most notable feature is a long ($>15^\circ$) and wide ($\sim 1^\circ$) arc with brightness temperatures T_B of 10–20 K, silhouetted against brighter H I. This arc delineates the northern rim of GSH139–03–69, a large ($18^\circ \times 10^\circ$) expanding H I supershell¹² (Fig. 2). The arc displays the strongest contrast with its surroundings where it crosses a bright 3.5° -diameter H I complex at Galactic coordinates $l = 136.5^\circ$, $b = +2.0^\circ$. (Here l and b are Galactic longitude and latitude, respectively.) The standard Galactic rotation model¹³ predicts that objects at the longitude and velocity (see Fig. 1) of GSH139–03–69 are 9 kpc from the Sun and 16 kpc from the Galactic Centre (1 kpc = 3.09×10^{21} cm). The arc is evident in the high-resolution, wide-field CGPS data, but has gone unnoticed in previous low-angular-resolution H I observations. However, once recognized in our data, it can be seen in those surveys^{12,14}.

We interpret the arc as the cold northern rim of GSH139–03–69 seen in self-absorption against bright, outer-Galaxy H I. That the arc is a long narrow region of space deficient in H I is unlikely, and its coincidence with GSH139–03–69 argues against its interpretation as a fortuitous arrangement of unrelated voids and/or small cold clouds. The arc is distinct in that it displays very little of the kinematic signature of large structures whose motions are dominated by Galactic rotation¹⁵ (manifested as an apparent shift in longitude with radial velocity). These distinctive kinematics are shared by the southern rim of GSH139–03–69.

More evidence that the arc is a cold object comes from its detection in H I absorption against unresolved extragalactic radio-continuum sources. Absorption spectra towards several bright ($T_B > 50$ K) 1.420-GHz continuum sources located behind the arc show higher line-centre optical depths τ_{max} and lower excitation ('spin') temperatures than do adjacent (within $\sim 1^\circ$) sources located behind bright H I (Fig. 3). (The spin temperature T_S of the H I is equal to its thermodynamic temperature in the interstellar medium.) Absorption against continuum sources yields only upper limits to the coldest temperatures encountered in the gas, and the effect of limited angular resolution is to underestimate optical depths. Unfortunately, we only found continuum sources

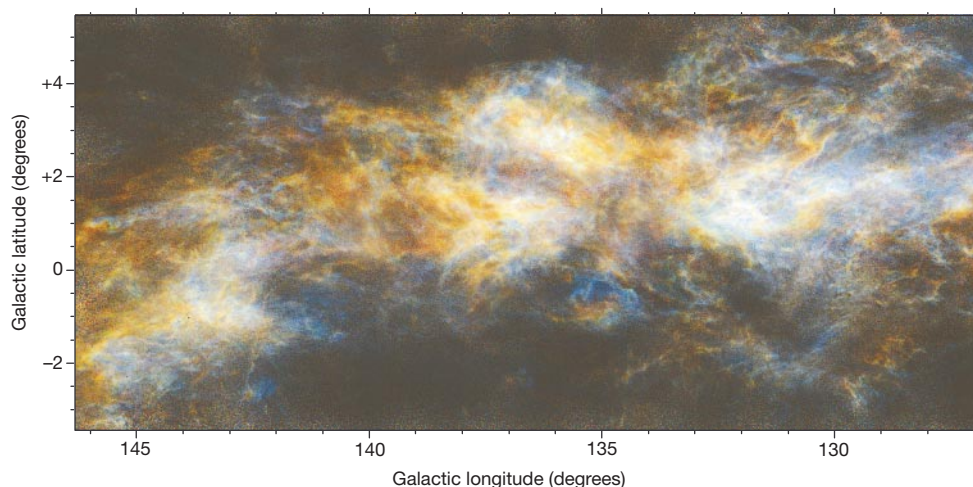


Figure 1 The cold H I arc as observed by the Canadian Galactic Plane Survey. The image covers a $19.5^\circ \times 8.9^\circ$ region towards the outer Galaxy; in Galactic coordinates (latitude b , longitude l), $126.8^\circ < l < 146.3^\circ$, $-3.5^\circ < b < +5.4^\circ$. Fifty-five overlapping interferometer fields were observed for this data set, and single-antenna data from the Low-Resolution DRAO Survey²⁶ were incorporated to produce a fully sampled spectroscopic mosaic with an angular resolution of $\sim 1'$ and a velocity resolution of

1.65 km s^{-1} . Red, H I brightness temperature with radial velocity -77.0 km s^{-1} ; green, -78.5 km s^{-1} ; blue, -80.0 km s^{-1} . Radial velocities are relative to the local standard of rest (LSR). The self-absorption arc, $\sim 1^\circ$ wide, is seen to stretch from $(l, b) \approx (131^\circ, -2^\circ)$ to $(146^\circ, +2^\circ)$, reaching its maximum latitude at $(l, b) \approx (139^\circ, +2^\circ)$. Brightness temperature scales linearly from 0 K (dark) to 40 K (bright). Assuming a distance of 9 kpc (see text), the linear scale of this image and of Fig. 2 is approximately 160 pc per degree.

sufficiently bright to make these measurements towards the fringes of the arc. However, the higher $\tau_{\max}$ and lower T_S observed towards the arc are evidence that it is composed of cold H I. We did not detect any absorption by the arc against continuum sources known to lie in the Perseus arm of our Galaxy (~ 2.5 kpc from the Sun), consistent with the arc being located more than 2.5 kpc from the Sun.

Analysis of the self-absorption yields direct measurements of the optical depth and temperature of the cold gas if—as is the case here—variations in the depth of the self-absorption can be measured as the background brightness varies. This type of analysis has rarely been applied, because self-absorption clouds have usually been identified only against uniform backgrounds. The arc has a mean (averaged over its central 5 km s^{-1} velocity range) T_B of about 20 K when silhouetted by the bright ($T_B = T_1 \approx 40$ K) cloud at $l = 136.5^\circ$ (Fig. 4), dropping to approximately 10 K against faint ($T_B = T_2 \approx 15$ K) backgrounds. Assuming that the temperature and optical depth are constant over the degree-scales over which we averaged data, we can derive them using a simple radiative transfer model: $T_1 \exp(-\tau_{\text{arc}}) + T_{\text{arc}}[1 - \exp(-\tau_{\text{arc}})] \approx 20$ K, and $T_2 \exp(-\tau_{\text{arc}}) + T_{\text{arc}}[1 - \exp(-\tau_{\text{arc}})] \approx 10$ K. We find that the arc has a very low temperature, $T_{\text{arc}} \approx 10$ K, and a high mean optical depth, $\tau_{\text{arc}} \approx 1$, values that are robust in the face of reasonable variations in the adopted average temperatures (Fig. 3). The typical $T_B = 5$ –10 K observed in the southern rim of GSH139–03–69 (in emission because it does not lie in front of bright H I at the same velocity) is consistent with our derived T_{arc} and τ_{arc} , and we conclude that they are broadly characteristic of the entire supershell. GSH139–03–69 contains $(1.9 \times 10^7) M_\odot$ of cold H I ($1 M_\odot$ is the solar mass, 1.99×10^{33} g). The presence of warm, optically thin H I lying in front of the arc would imply that temperatures in the cold gas are even lower than we have measured: the requirement that the arc can be no colder than the cosmic microwave background value of 2.7 K implies the warm H I column density is $< 5 \times 10^{19} \text{ cm}^{-2}$. The arc is not detected in CO surveys^{21,16}, which implies an H₂ column density that is less than $5 \times 10^{19} \text{ cm}^{-2}$. A more significant limit comes from considering the balance between the dissociation and formation rate¹⁷ of H₂: given the H I column density towards the

arc ($1.1 \times 10^{20} \text{ cm}^{-2}$) and its inferred density (1.5 cm^{-3}), the H₂ abundance must be very low, much less than 1% by mass. We conclude the arc is composed overwhelmingly of cold atomic hydrogen.

H I at temperatures as low as 10 K is rarely observed, and then only in compact (< 10 pc), low-mass ($< 10^2 M_\odot$) clouds³, and in a few somewhat larger ($\sim 10^2$ pc, $\sim 10^3 M_\odot$) clouds in the solar neighbourhood^{4,5}. Molecular clouds have envelopes of H I, but they are generally warmer¹⁸ than the interior H₂ and contain only about half the mass of the entire cloud¹⁹. GSH139–03–69, of size $2.8 \text{ kpc} \times 1.6 \text{ kpc}$, is comparable in mass to the largest ($\sim 10^2$ pc, $\sim 10^6 M_\odot$) giant molecular cloud complexes⁶, but is equally cold. It is also colder than the massive ($\sim 10^7 M_\odot$) H I superclouds²⁰. Such a large and massive structure composed solely of cold H I has not previously been recognized in the Galaxy.

The bulk of the mass of supershells is thought to be in a shell of swept-up interstellar gas at the surface of a gigantic, roughly spherical, shock wave. The compressed post-shock gas cools radiatively: our results show it can reach lower temperatures than the upper limits (30–300 K) for interstellar shells derived from H I absorption against continuum sources^{21,22}. Although the mechanisms that create supershells remain obscure, they play a role in models of the multi-phase interstellar medium in which material for new clouds is accumulated in gas swept up by supernovae and possibly other explosive processes^{23,24}. The recognition that supershells can be cold advances our knowledge of the properties of shells,

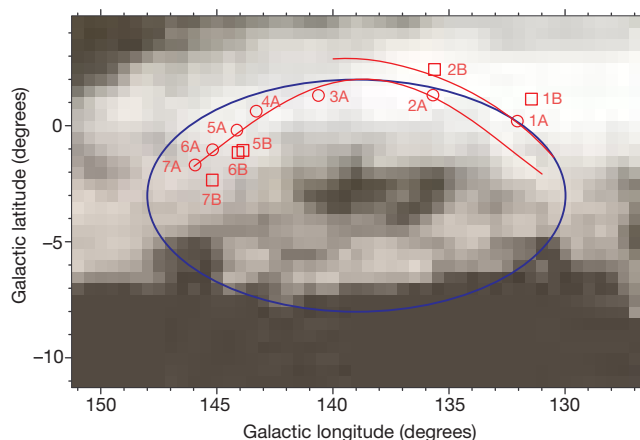


Figure 2 The H I supershell GSH139–03–69. These data are taken from ref. 14. The grey scale is mean brightness temperature (T_B in K) of the H I line averaged over the velocity range (-87 to -59 km s^{-1}) of the supershell¹²; T_B scales logarithmically from 0 K (dark) to 30 K (bright). The pixel size (0.5°) is approximately equal to the angular resolution of the data. The approximate outline of the supershell as originally proposed¹² is indicated by the blue ellipse: the data here suggest that the southern rim of the supershell in fact lies farther north. The two red curves at the northern rim of the ellipse indicate the approximate location of the self-absorption arc. The northern edge of the arc is not well defined at longitudes in excess of 140° . Red circles, positions of 'on-arc' absorption spectra against radio-continuum sources; red squares, positions of the 'off-arc' spectra (see Fig. 3).

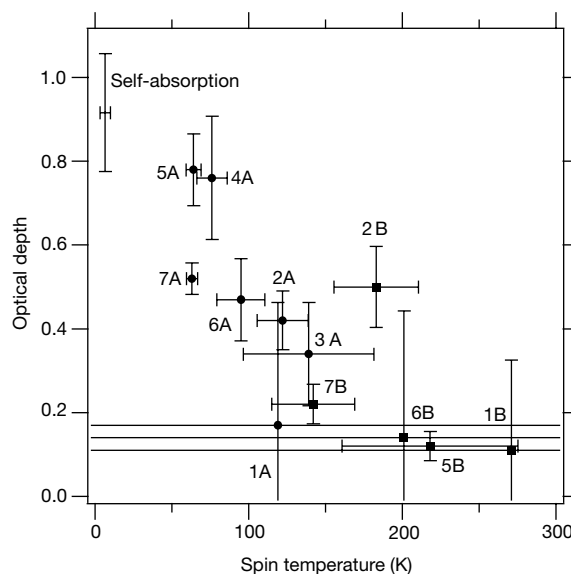


Figure 3 Evidence for cold H I towards the dark arc from continuum absorption. Plotted is line centre optical depth $\tau_{\max}$ versus spin temperature T_S for H I line absorption against 1.420-GHz continuum sources. Spectra were derived for seven sources towards the fringes of the arc (circles, labelled with the suffix A) and five towards bright background emission (squares, labelled with the suffix B). Pairs of on-arc and off-arc sources that lie within $\sim 1^\circ$ on the sky of each other are labelled with the same numerical prefix. Source identifications: 1A, KR137; 2A, F3R3994; 3A, KR154; 4A, KR164; 5A, KR168; 6A, F3R4233; 7A, KR179; 1B, KR135; 2B, F3R3992; 5B, 3C86A; 6B, F3R4198; 7B, KR175. The positions of these sources are shown in Fig. 2. Error bars were calculated using the noise statistics of the data: it is characteristic of absorption spectroscopy that the uncertainty in the derived T_S can be very large when the $\tau_{\max}$ and/or the uncertainty in $\tau_{\max}$ is large. The point labelled 'Self-absorption' marks the $\tau_{\max}$ and T_S for the arc derived from analysis of the self-absorption, with error bars representing the effect of a 5% uncertainty in the measured average T_B . The on-arc spectra have generally higher $\tau_{\max}$ (> 0.3) and lower T_S (< 140 K) than do the off-arc spectra ($\tau_{\max} < 0.3$ and $T_S > 140$ K), indicating that lines of sight towards the arc contain substantial quantities of gas at low temperatures.

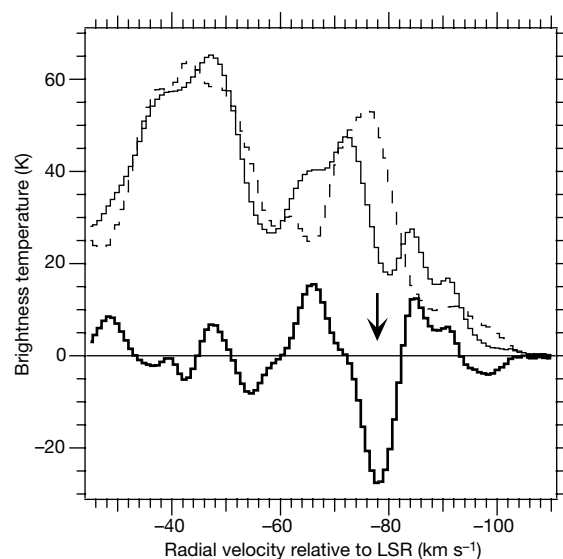


Figure 4 On-arc and off-arc H I spectra. The spectra were produced by averaging in latitude-longitude boxes that were centred towards the arc at $l = 137.0^\circ$, $b = +2.1^\circ$ (thin black line) and towards two off-arc positions (dashed line) offset symmetrically to either side of the on-arc position. The on-arc minus off-arc spectrum is shown by the thick black line. The arrow indicates the velocity of the deepest self-absorption. Conventional spectra, covering a wide range in velocity (in which an integrating or averaging box has a fixed angular size), suffer from strongly velocity-dependent linear scales owing to Galactic rotation. Thus, the physical area of space in the plane of the sky sampled by a conventional box is smaller at low distances (in the outer Galaxy, corresponding to less negative velocities) and larger at large distances (more negative velocities), which makes

physical interpretation difficult. In order to maintain a constant box area in these spectra, we scaled the angular size of the averaging box at each velocity according to the distance-velocity relation derived from the Galactic rotation curve¹³. At the velocity of the arc, the box areas were 0.16 degree^2 (area, about $4,000 \text{ pc}^2$). The box centres were also shifted with velocity to maintain a constant distance in the plane of the sky of 170 pc between the on-arc and off-arc boxes. This technique does not maintain a constant box volume with distance: this would require data with impracticably high velocity resolution. To our knowledge this box scaling technique has not before been applied to spectra of the Galactic interstellar medium.

their fragmentation into clouds, and the role of cold H I in the Galaxy.

The conventional wisdom that cold, massive clouds in the interstellar medium are always predominantly molecular has achieved near-axiomatic status. The discovery that H I supershells can contain some of the coldest gas in the Galaxy will require a re-evaluation of this view. We have shown that Galactic structures made of cold H I can be much more massive than previously thought: four decades in mass separate GSH139-03-69 from classical clouds of cold H I. The properties of such clouds that might exist between these two extremes of mass are at present unknown. The expected fragmentation of supershells through gravitational or other types of instability²⁵ is one possible source of as-yet unobserved clouds of cold H I in the range $(10^4\text{--}10^6)M_\odot$. The advent of wide-field, high-angular-resolution, H I surveys—such as the Canadian Galactic Plane Survey—offers us the opportunity to examine in more detail the occurrence of cold H I in our Galaxy. □

Received 5 April; accepted 16 May 2001.

- Kulkarni, S. R. & Heiles, C. in *Interstellar Processes* (eds Hollenbach, D. J. & Thronson, H. A.) 87–122 (Reidel, Dordrecht, 1987).
- Dame, T. M., Hartmann, D. & Thaddeus, P. The Milky Way in molecular clouds: A new composite CO survey. *Astrophys. J.* **547**, 792–813 (2001).
- Gibson, S. J., Taylor, A. R., Higgs, L. A. & Dewdney, P. E. A new view of cold H I clouds in the Milky Way. *Astrophys. J.* **540**, 851–862 (2000).
- Trapero, J. *et al.* A cold massive interstellar cloud within 120 parsecs of the Sun: K I optical and H I radio observations. *Astrophys. J.* **445**, 231–239 (1995).
- Crutcher, R. M. & Riegel, K. W. Optical interstellar line studies of a nearby cold cloud. *Astrophys. J.* **188**, 481–488 (1974).
- Solomon, P. M., Rivolo, A. R., Barrett, J. & Yahil, A. Mass, luminosity, and line width relations of Galactic molecular clouds. *Astrophys. J.* **319**, 730–741 (1987).
- Dickey, J. M. & Lockman, F. J. H I in the Galaxy. *Annu. Rev. Astron. Astrophys.* **28**, 215–261 (1990).
- Spitzer, L. *Physical Processes in the Interstellar Medium* Ch. 3 (Wiley & Sons, New York, 1978).
- Landecker, T. L. *et al.* The synthesis telescope at the Dominion Radio Astrophysical Observatory. *Astron. Astrophys. Suppl. Ser.* **145**, 509–524 (2000).
- Higgs, L. A. in *New Perspectives on the Interstellar Medium* (eds Taylor, A. R., Landecker, T. L. & Joncas, G.) 15–26 (Astronomical Society of the Pacific, San Francisco, 1999).
- Taylor, A. R. *et al.* The Canadian Galactic Plane Survey. *Astron. J.* (submitted).
- Heiles, C. H I shells and supershells. *Astrophys. J.* **229**, 533–544 (1979).
- Kerr, F. J. & Lynden-Bell, D. Review of Galactic constants. *Mon. Not. R. Astron. Soc.* **221**, 1023–1038 (1986).
- Hartmann, D. & Burton, W. B. *Atlas of Galactic Neutral Hydrogen* (Cambridge Univ. Press, Cambridge, 1997).
- Burton, W. B. in *Galactic and Extragalactic Radio Astronomy* (eds Verschuur, G. L. & Kellermann, K. I.) 295–358 (Springer, Berlin, 1988).
- Heyer, M. H. *et al.* The Five College Radio Astronomy Observatory CO survey of the outer Galaxy. *Astrophys. J. Suppl. Ser.* **115**, 241–258 (1998).
- Van Dishoeck, E. F. & Black, J. H. Comprehensive models of diffuse interstellar clouds: Physical conditions and molecular abundances. *Astrophys. J. Suppl. Ser.* **62**, 109–145 (1986).
- Wannier, P. G., Lichten, S. M. & Morris, M. Warm H I halos around molecular clouds. *Astrophys. J.* **268**, 727–738 (1983).
- Blitz, L. in *The Physics of Star Formation and Early Stellar Evolution* (eds Lada, C. J. & Kylafis, N. D.) 3–33 (Kluwer Academic, Dordrecht, 1991).
- Elmegreen, B. G. & Elmegreen, D. M. H I superclouds in the inner Galaxy. *Astrophys. J.* **320**, 182–198 (1987).
- Kulkarni, S. R., Dickey, J. M. & Heiles, C. Measurement of spin temperatures in a rapidly moving H I shell. *Astrophys. J.* **291**, 716–721 (1985).
- Heiles, C. Temperatures and their variation within interstellar H I structures. *Astrophys. J.* **262**, 135–141 (1982).
- Perna, R. & Raymond, J. Testing the multiple supernovae versus gamma-ray burst scenarios for giant H I supershells. *Astrophys. J.* **539**, 706–711 (2000).
- Tenorio-Tagle, G. & Bodenheimer, P. Large-scale expanding superstructures in galaxies. *Annu. Rev. Astron. Astrophys.* **26**, 145–197 (1988).
- Ehlerová, S., Palouš, J., Theis, Ch. & Hensler, G. Fragmentation of expanding shells in spiral and irregular galaxies. *Astron. Astrophys.* **328**, 121–129 (1997).
- Higgs, L. A. & Tapping, K. F. The Low-Resolution DRAO Survey of H I emission from the Galactic plane. *Astron. J.* **120**, 2471–2487 (2000).

Acknowledgements

We thank T. Dame for providing the CO data from ref. 2. The Canadian Galactic Plane Survey (CGPS) is a Canadian project with international partners. The Dominion Radio Astrophysical Observatory is operated as a national facility by the National Research Council of Canada. The CGPS is supported by the Natural Sciences and Engineering Research Council of Canada.

Correspondence and requests for materials should be addressed to L.B.G.K. (e-mail: lewis.knee@nrc.ca).

Isotopic homogeneity of iron in the early solar nebula

X. K. Zhu, Y. Guo, R. K. O'Nions, E. D. Young & R. D. Ash

Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK

The chemical and isotopic homogeneity of the early solar nebula, and the processes producing fractionation during its evolution, are central issues of cosmochemistry. Studies of the relative abundance variations of three or more isotopes of an element can in principle determine if the initial reservoir of material was a homogeneous mixture or if it contained several distinct sources of precursor material. For example, widespread anomalies^{1–4} observed in the oxygen isotopes of meteorites have been interpreted as resulting from the mixing of a solid phase that was enriched in ¹⁶O with a gas phase in which ¹⁶O was depleted^{1–3}, or as an isotopic 'memory' of Galactic evolution⁵. In either case, these anomalies are regarded as strong evidence that the early solar nebula was not initially homogeneous. Here we present measurements of the relative abundances of three iron isotopes in meteoritic and terrestrial samples. We show that significant variations of iron isotopes exist in both terrestrial and extraterrestrial materials. But when plotted in a three-isotope diagram, all of the data for these Solar System materials fall on a single mass-fractionation line, showing that homogenization of iron isotopes occurred in the solar nebula before both planetesimal accretion and chondrule formation.

The significance of the oxygen-isotope anomalies observed in meteorites with regard to early Solar System history remains the subject of active debate. It has been demonstrated experimentally that the observed anomalies could be produced by chemical reactions within a homogeneous gaseous reservoir^{6,7}. Indeed, it has been observed that mass-independent fractionation of oxygen isotopes is widespread in the Earth's atmosphere^{7–9}. More recent studies have shown that such oxygen-isotope anomalies induced by chemical reactions between gas phases can be transferred to terrestrial solid phases^{10,11}. Thus the existence of oxygen-isotope anomalies can no longer be regarded as evidence of the heterogeneity of the early solar nebula.

To constrain issues such as the chemical and isotopic homogeneity in the early solar nebula, investigation of the isotopic composition of elements other than oxygen is therefore needed. Thanks to recent advances in mass spectrometry, the opportunity now exists to investigate the patterns of isotopic variation in a range of other elements of different volatility. Among these elements, iron is a particularly interesting target. It is a major element in terrestrial planets, and has four naturally occurring stable isotopes, ⁵⁴Fe, ⁵⁶Fe, ⁵⁷Fe and ⁵⁸Fe. These isotopes are produced at the explosive silicon-burning stage of stellar evolution¹². Isotopes of iron are produced in different proportions in different stars, and the average isotopic composition of the interstellar medium changes with time during Galactic evolution (D. D. Clayton, at <http://photon.phys.clemson.edu/joel/claytonZ/>). Variation patterns of iron isotopes resulting from nucleosynthetic processes are significantly different from those caused by mass-dependent fractionation (see above). Thus iron isotopes are expected to be useful tracers in constraining the degree of initial isotopic homogeneity of the solar nebula. Realization of this potential has been largely hampered by the absence of suitable techniques for precise isotope measurements. The introduction of plasma source mass spectrometry is rapidly advancing the techniques of isotope analysis^{13–17}, and a technique for high-precision measurement of the ⁵⁷Fe/⁵⁴Fe ratio has been developed recently^{15,16}. Here the technique is developed further, and we

present the results of measurements of the ⁵⁶Fe/⁵⁴Fe and ⁵⁷Fe/⁵⁴Fe ratios in terrestrial and extraterrestrial materials.

We have sampled a wide variety of meteorite types, representing different parent bodies. These include chondrites (carbonaceous, ordinary, and enstatite chondrites), achondrites (aubrite, eucrite, ureilite, and Shergotty–Nakhla–Chassigny (SNC) meteorite), pallasites, and iron meteorites. We performed measurements of iron isotopes on chondrules, matrices, metal fractions, or bulk samples. Chondrules were hand-picked from the meteorites after gentle crushing, and freed from matrix coating by air abrasion and ultrasonic cleaning. For reference purposes, we also analysed a selection of iron-containing minerals—haematite, magnetite and siderite. Silicate meteoritic materials were digested in concentrated HF and HNO₃, iron meteorites in 2 M HCl, and terrestrial minerals in 6 M HCl. All samples were purified by ion exchange separation following complete digestion¹⁵. Iron-isotope analyses were performed on a Nu Instruments plasma source mass spectrometer, using a standard–sample–standard procedure similar to that described previously^{14–16}. The isobaric interferences of [¹⁴N⁴⁰Ar]⁺ and [¹⁶O⁴⁰Ar]⁺ with the ⁵⁴Fe and ⁵⁶Fe signals were reduced to a negligible levels. This was done by matching concentrations of Fe in standard and sample solutions to within 20%, and by desolvating samples using a CETAC MCN-6000 nebulizer without N₂ flow. Measured ⁵⁶Fe/⁵⁴Fe and ⁵⁷Fe/⁵⁴Fe ratios are expressed in terms of $\epsilon^{56}\text{Fe}$ and $\epsilon^{57}\text{Fe}$, which are deviations in parts per 10⁴ from the Fe isotope reference material IRMM-14. Results are presented in Table 1.

These materials show overall variations of ~16 and ~23 ϵ -units in ⁵⁶Fe/⁵⁴Fe and ⁵⁷Fe/⁵⁴Fe ratios, respectively, with $\epsilon^{56}\text{Fe}$ ranging from –6.1 to 9.8, and $\epsilon^{57}\text{Fe}$ varying from –8.8 to 14.1 (Table 1). Within the meteoritic samples, the chondritic materials show relatively large variations with $\epsilon^{57}\text{Fe}$ ranging from –8.8 to 9.3, whereas achondrites and pallasites display much smaller variations with $-2.1 \leq \epsilon^{57}\text{Fe} \leq 2.0$, and iron meteorites are particularly homogeneous with $\epsilon^{57}\text{Fe}$ varying from 0.6 to 2.0 only. These results for iron meteorites are indistinguishable from the data reported previously for the same samples¹⁶, where only ⁵⁷Fe/⁵⁴Fe ratios had been measured. The consistency between these two data sets shows the robustness of the analytical techniques used. The smaller variation of iron isotopes in achondrites, pallasites and iron meteorites compared to chondritic materials could be due to isotope re-equilibration having occurred within the parent bodies.

When plotted on a three-isotope diagram, all iron-isotope data

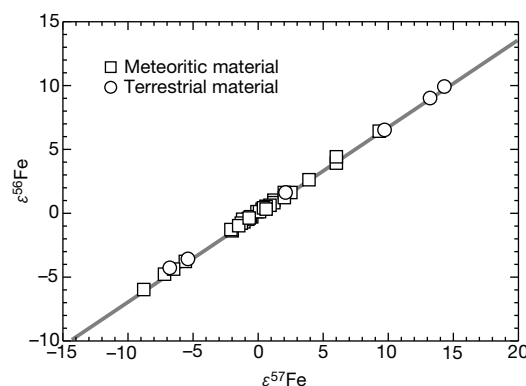


Figure 1 A three-isotope plot for iron, showing that all the Fe-isotope data from both terrestrial and extraterrestrial materials plot on a single mass-fractionation line. The correlation between $\epsilon^{56}\text{Fe}$ and $\epsilon^{57}\text{Fe}$ is $\epsilon^{56}\text{Fe} = (0.677 \pm 0.004)\epsilon^{57}\text{Fe}$, and $r^2 = 0.999$. If the regression is not forced through the origin, then $\epsilon^{56}\text{Fe} = (0.678 \pm 0.004)\epsilon^{57}\text{Fe} - 0.034$, with $r^2 = 0.999$. The size of data points represents the analytical uncertainty, which is defined as two standard deviations of long-term repeatability.

define a single line that passes through the origin with a slope of ~ 0.68 (Fig. 1). This observation is entirely consistent with theoretical expectations if the iron-isotope compositions in all analysed materials evolved from a single isotopically uniform reservoir in a mass-dependent manner. As the samples analysed are ultimately derived from different parent bodies, the iron-isotope homogenization cannot be attributed to processes occurring within the parent bodies themselves. Thus the observation of this single mass-fractionation line shows that all the iron in both terrestrial and extraterrestrial materials analysed were derived ultimately from a single isotopically homogeneous reservoir—which existed before the formation of chondrules and parent bodies.

Extraterrestrial samples analysed in this study cover all the main known meteorite types, and probably represent 14 different parent bodies including the Earth and Mars. Thus there is good reason to believe that the iron isotopes recorded in these samples are representative of Solar System materials. Furthermore, the measured samples include meteorites that are amongst the most anomalous materials from the oxygen isotope point of view, and they exhibit

the full range in oxygen isotope ratios observed for bulk meteorites. For example, the meteorite with the most ^{16}O -rich bulk composition is the pallasite Eagle Station, with $\Delta^{17}\text{O} = -4.51\text{‰}$ (refs 18, 19); $\Delta^{17}\text{O}$ are deviations of $^{17}\text{O}/^{16}\text{O}$ ratios from the terrestrial mass fractionation line. The most depleted is LL3 chondrite Chainpur, with $\Delta^{17}\text{O} = +1.04\text{‰}$ (ref. 20). Furthermore, we have measured individual chondrules from both the carbonaceous chondrite Allende and the ordinary chondrite Chainpur for Fe isotopes. It has been documented³ that chondrules from carbonaceous chondrites and ordinary chondrites are distinctively different in oxygen-isotope features, and it has been argued that these two sets of chondrules sampled the solar nebula in different regions or at different times³. So we expect that if any iron-isotope anomalies were inherited from circumstellar or interstellar mediums by the parent-body materials, they would probably have been revealed within this sample suite.

As a moderately volatile element, iron is depleted in refractory presolar grains such as corundum (Al_2O_3), and no attempt has been made so far to assess these phases for possible presolar iron-isotope signatures. We note, however, that isotopic anomalies of iron—as well as nickel and chromium, which have similar volatilities to iron—have been claimed to exist in some calcium- and aluminium-rich refractory inclusions (CAIs)^{21–23}. Also, chromium-isotope anomalies have been found in leachates of some meteorites^{24,25}. This demonstrates that iron isotopes were unlikely to have been initially homogeneous, and that their homogeneity recorded in the Solar System material results from subsequent processes taking place within the solar nebula.

This study shows that the solar nebula probably underwent events of isotope homogenization at an early stage, then subsequently fractionated in a mass-dependent manner for metal isotopes before (or during) the formation of chondrules and the parent bodies of meteorites. The homogeneity of iron isotopes in the early solar nebula is in marked contrast to the widespread oxygen-isotope anomalies recorded in meteorites^{1–4}. But this apparent conflict could be reconciled if the latter anomalies resulted predominantly from chemical reactions that took place in the solar nebula, or if complete homogenization for oxygen isotopes had never been achieved between solid materials and the gaseous reservoir in the solar nebula. Because isotopic anomalies of Ti, Ca, Si, Cr and Fe have been reported in some CAIs^{2,21–23,26–28}, the isotope-homogenization events are unlikely to have pre-dated CAI formation, but must have preceded the formation of chondrules and the parent bodies of meteorites. □

Table 1 Fe-isotope data of terrestrial and extraterrestrial materials

Sample	Description	$\epsilon^{57}\text{Fe}$	$\epsilon^{56}\text{Fe}$
Carbonaceous Chondrites			
Allende (CV3)	Bulk sample (AG22)	-0.5	-0.4
Allende (CV3)	Matrix (AG23)	1.2	0.9
Allende (CV3)	Chondrule (AG38)	-8.8	-6.1
Allende (CV3)	Chondrule (A1)	-5.6	-3.9
Allende (CV3)	Chondrule (A2)	-7.2	-4.9
Allende (CV3)	Chondrule (A3)	3.9	2.5
Allende (CV3)	Chondrule (A6)	9.3	6.3
Murchison (CM2)	Bulk sample (BM5484)	-0.1	0.0
Orgueil (CI)	Bulk sample (BM3673)	6.0	3.8
Ordinary chondrites			
Chainpur (LL3)	Matrix (C19)	6.0	4.3
Chainpur (LL3)	Chondrule (C12)	-6.5	-4.6
Chainpur (LL3)	Chondrule (C14)	0.3	0.2
Chandakapur (L5)	Bulk sample (Mt4C)	-2.0	-1.5
Forest City (H5)	Metal fraction (Mt180)	-0.8	-0.6
Forest City (H5)	Duplicate	-0.6	-0.4
Enstatite chondrite			
Indarch (EH4)	Bulk sample	-1.1	-0.8
Achondrites			
Khor Temiki	Aubrite, bulk sample	0.1	0.0
Dyalpur	Ureilite, bulk sample	2.5	1.5
Stannern	Eucrite, bulk sample (Mt47)	-1.2	-0.6
Nakhla	Martian meteorite, bulk (Mt84)	-2.1	-1.4
Pallasites			
Krasnojarsk	Main group pallasite, metal fraction (Mt1)	-1.3	-0.9
Krasnojarsk	Duplicate	-1.5	-1.1
Ilmāc	Main group pallasite, metal fraction (Mt77)	-0.7	-0.5
Eagle Station	Non-main group pallasite, metal fraction	1.2	0.7
Iron meteorites			
Toluca (IAB)	Bulk sample (Mt70)	0.6	0.4
Cranbourne (IAB)	Bulk sample (Mt99)	2.0	1.5
Coahuila (IAB)	Bulk sample (Mt83)	0.9	0.5
Carthage (IIAB)	Bulk sample (Mt42)	0.6	0.4
Thunda (IIAB)	Bulk sample (Mt46)	0.4	0.3
Bella Rocca (IIIAB)	Bulk sample (Mt41)	0.6	0.2
Tazewell (IIICD)	Bulk sample (Mt33)	2.0	1.1
Terrestrial material			
Haematite	Specular; Cleator Moor, UK (OUM1927)	13.2	8.9
Haematite	Massive; Lake Superior (OUM858)	14.3	9.8
Haematite	Specular; Mt Vesuvius, Italy (OUM11387)	-5.4	-3.7
Magnetite	Rio La Marra Elba, Italy (OUM11421)	9.7	6.4
Magnetite	Kolar Peninsular, Russia (OUM29789)	2.1	1.5
Siderite	Newdorf, Harz Mt, Germany (OUM11479)	-6.8	-4.4

Measured $^{57}\text{Fe}/^{56}\text{Fe}$ and $^{56}\text{Fe}/^{54}\text{Fe}$ ratios are expressed as $\epsilon^{57}\text{Fe}$ and $\epsilon^{56}\text{Fe}$ which are deviations in parts per 10⁴ from the Fe-isotope reference material IRMM-14 as following: $\epsilon^i\text{Fe} = [(R_{\text{sample}}/R_{\text{standard}}) - 1] \times 10,000$, where $R = {}^i\text{Fe}/^{56}\text{Fe}$. Values of $\epsilon^{57}\text{Fe}$ and $\epsilon^{56}\text{Fe}$ for each sample are the averages of two or three separate analyses. Duplicate analyses include duplication of ion exchange chromatographic separation. The data quality was also controlled by repeated measurements of Romil Fe solution versus IRMM-14 Fe-isotope reference material during the run. The long-term repeatability for both ^{57}Fe and ^{56}Fe is 0.6‰ at the 2-standard-deviation level, which gives an external precision of 0.6‰ at the 2 σ level. The internal precision for individual analysis is generally better than 0.3‰ at the 2 σ level. Symbols in brackets of the first column are meteorite types, and those in brackets of the second column are sample numbers.

Received 13 October 2000; accepted 15 May 2001.

1. Clayton, R. N., Grossman, L. & Mayeda, T. K. A component of primitive nuclear composition in carbonaceous meteorites. *Science* **182**, 485–488 (1973).
2. Clayton, R. N., Hinton, R. W. & Davis, A. M. Isotopic variations in the rock-forming elements in meteorites. *Phil. Trans. R. Soc. Lond. A* **325**, 483–501 (1988).
3. Clayton, R. N. Oxygen isotopes in meteorites. *Annu. Rev. Earth Planet. Sci.* **21**, 115–149 (1993).
4. Young, E. D. & Russell, S. S. Oxygen reservoirs in the early solar nebula inferred from an Allende CAI. *Science* **282**, 452–455 (1998).
5. Clayton, D. D. Isotope anomalies: Chemical memory of galactic evolution. *Astrophys. J.* **334**, 191–195 (1988).
6. Thieme, M. H. & Heidenreich, J. E. The mass independent fractionation of oxygen: A novel isotope effect and its possible cosmochemical implications. *Science* **219**, 1073–1075 (1983).
7. Thieme, M. H. Atmosphere science—Mass-independent isotope effects in planetary atmospheres and the early solar system. *Science* **283**, 341–345 (1999).
8. Cliff, S. S. & Thieme, M. H. The $^{18}\text{O}/^{16}\text{O}$ and $^{17}\text{O}/^{16}\text{O}$ ratios in atmospheric nitrous oxide: A mass-independent anomaly. *Science* **278**, 1774–1776 (1997).
9. Rockmann, T. et al. Mass-independent oxygen isotope fractionation in atmospheric CO as a result of the reaction CO + OH. *Science* **281**, 544–546 (1998).
10. Bao, H. et al. Anomalous ^{17}O compositions in massive sulphate deposits on the Earth. *Nature* **406**, 176–178 (2000).
11. Bao, H., Campbell, D. A., Bockheim, J. G. & Thieme, M. H. Origin of sulphate in Antarctic dry-valley soils as deduced from anomalous ^{17}O compositions. *Nature* **407**, 499–502 (2000).
12. Arnett, D. *Supernovae and Nucleosynthesis* (Princeton Univ. Press, Princeton, 1996).
13. Halliday, A. N. et al. Applications of multiple collector ICPMS to cosmochemistry, geochemistry and paleoceanography. *Geochim. Cosmochim. Acta* **62**, 919–940 (1998).
14. Zhu, X. K., O'Nions, R. K., Guo, Y., Belshaw, N. S. & Rickard, D. Determination of natural Cu-isotope variation by plasma-source mass spectrometry: implications for use as geochemical tracers. *Chem. Geol.* **163**, 139–149 (2000).

15. Zhu, X. K., O'Nions, R. K., Guo, Y. & Reynolds, B. C. Secular variation of iron isotopes in North Atlantic deep water. *Science* **287**, 2000–2002 (2000).
16. Belshaw, N. S., Zhu, X. K., Guo, Y. & O'Nions, R. K. High precision measurement of iron isotopes by plasma source mass spectrometry. *Int. J. Mass. Spectrom.* **197**, 191–195 (2000).
17. Marechal, C. N., Telouk, P. & Albarède, F. Precise analysis of copper and zinc isotopic compositions by plasma-source mass spectrometry. *Chem. Geol.* **156**, 251–273 (1999).
18. Clayton, R. N. & Mayeda, T. K. Formation of ureilites by nebular processes. *Geochim. Cosmochim. Acta* **52**, 1313–1318 (1988).
19. Clayton, R. N. & Mayeda, T. K. Oxygen isotope studies of achondrites. *Geochim. Cosmochim. Acta* **60**, 1999–2017 (1996).
20. Clayton, R. N. Mayeda, T. K., Goswami, J. N. & Olsen, E. J. Oxygen isotope studies of ordinary chondrites. *Geochim. Cosmochim. Acta* **55**, 2317–2337 (1991).
21. Volkening, J. & Papanastassiou, D. A. Iron isotope anomaly. *Astrophys. J.* **347**, L43–L46 (1989).
22. Volkening, J. & Papanastassiou, D. A. Zinc isotope anomaly. *Astrophys. J.* **358**, L29–L32 (1990).
23. Birck, J. L. & Lugmair, G. W. Nickel and chromium isotopes in Allende inclusions. *Earth Planet. Sci. Lett.* **90**, 131–143 (1988).
24. Rotaru, M., Birck, J. L. & Allègre, C. J. Clue to early solar system history from chromium isotopes in carbonaceous chondrites. *Nature* **358**, 465–470 (1992).
25. Podošek, F. A. et al. Thoroughly anomalous chromium in Orgueil. *Meteorit. Planet. Sci.* **32**, 617–627 (1997).
26. Niemeier, S. & Lugmair, G. W. Titanium isotopic anomalies in meteorites. *Geochim. Cosmochim. Acta* **48**, 1401–1416 (1984).
27. Niemeier, S. & Lugmair, G. W. Ubiquitous isotopic anomalies in Ti from normal Allende inclusions. *Earth Planet. Sci. Lett.* **53**, 211–225 (1981).
28. Papanastassiou, D. A. Chromium isotopic anomalies in the Allende meteorite. *Astrophys. J.* **308**, L27–L30 (1986).

Acknowledgements

We thank N. S. Belshaw for assistance with mass spectrometry; A. Galy for discussions and for providing solution aliquots of some samples analysed; G. Turner and F. Albarède for comments on the manuscript; and M. Price for providing samples of the Mt and OUM series. This work was supported by the Natural Environment Research Council.

Correspondence and requests for materials should be addressed to X.K.Z.
(e-mail: xiangz@earth.ox.ac.uk).

Entanglement of the orbital angular momentum states of photons

Alois Mair*, Alipasha Vaziri, Gregor Weihs & Anton Zeilinger

Institut für Experimentalphysik, Universität Wien, Boltzmanngasse 5, 1090 Wien, Austria

Entangled quantum states are not separable, regardless of the spatial separation of their components. This is a manifestation of an aspect of quantum mechanics known as quantum non-locality^{1,2}. An important consequence of this is that the measurement of the state of one particle in a two-particle entangled state defines the state of the second particle instantaneously, whereas neither particle possesses its own well-defined state before the measurement. Experimental realizations of entanglement have hitherto been restricted to two-state quantum systems^{3–6}, involving, for example, the two orthogonal polarization states of photons. Here we demonstrate entanglement involving the spatial modes of the electromagnetic field carrying orbital angular momentum. As these modes can be used to define an infinitely dimensional discrete Hilbert space, this approach provides a practical route to entanglement that involves many orthogonal quantum states, rather than just two. Multi-dimensional entangled states could be of considerable importance in the field of quantum information^{7,8}, enabling, for example, more efficient use of communication channels in quantum cryptography^{9–11}.

Multi-dimensional entanglement is a way—in addition to multi-particle entanglement—to extend the usual two-dimensional two-particle state. There have been suggestions^{12,13} (and only a proof-of-

principle experiment¹⁴) as to how to realize higher-order entanglement via multiport beam splitters. Here we present an experiment in which we used the spatial modes of the electromagnetic field carrying orbital angular momentum to create multi-dimensional entanglement. The advantage of using these modes to create entanglement is that they can be used to define an infinitely dimensional discrete (because of the quantization of angular momentum) Hilbert space.

The experimental realization proceeded in the following two steps. First, we confirmed that spontaneous parametric down-conversion conserves the orbital angular momentum of photons. This was done for pump beams carrying orbital angular momenta of $-\hbar$, 0 and $+\hbar$ per photon, respectively. Second, we showed that the state of the down-converted photons can not be explained by assuming classical correlation—in the sense that the photon pairs produced are just a mixture of the combinations allowed by conservation of angular momentum. We proved that, in contrast, they are a coherent superposition of these combinations, and hence they have to be considered as entangled in their orbital angular momentum. After completion of the experimental work presented here, related theoretical work was brought to our attention^{15,16}.

We will now discuss in order the two steps mentioned above. For paraxial light beams, Laguerre–gaussian (LG) modes (Fig. 1) define a possible set of basis vectors. As predicted¹⁷ and observed¹⁸, LG modes carry an orbital angular momentum for linearly polarized light that is distinct from the intrinsic angular momentum of photons associated with their polarizations. This external angular momentum of the photon states is the reason why they have been suggested for gearing micromachines, and it has been shown that they can be used as optical tweezers^{19–21}.

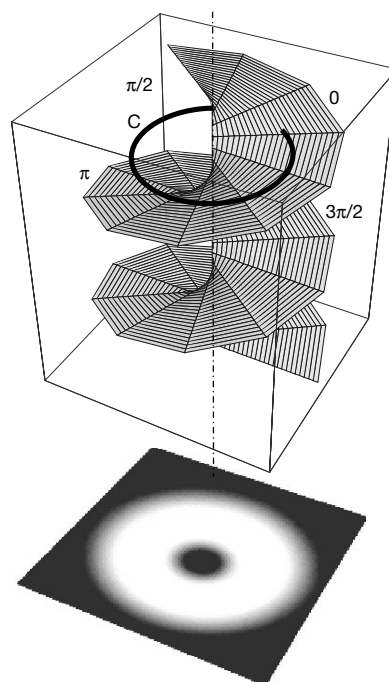


Figure 1 The wave front (top) and the intensity pattern (bottom) of the simplest Laguerre–gaussian (LG_0) or ‘doughnut’ mode. The index l is referred to as the winding number, and $(p + 1)$ is the number of radial nodes. Here we only consider cases of $p = 0$. The customary gaussian mode can be viewed as an LG mode with $l = 0$. The handedness of the helical wave fronts of the LG modes is linked to the sign of the index l , and can be chosen by convention. The azimuthal phase term $\exp i l \phi$ of the LG modes results in helical wave fronts. The phase variation along a closed path C around the beam centre is $2\pi l$. Therefore, in order to fulfil the wave equation, the intensity has to vanish in the centre of the beam.

* Present address: Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

To demonstrate the conservation of the orbital angular momentum carried by the LG modes in spontaneous parametric down-conversion, we investigated three different cases—for pump photons possessing orbital angular momenta of $-\hbar$, 0 and $+\hbar$ per photon, respectively. As a pump beam, we used an argon-ion laser (wavelength 351 nm) which we could operate either with a simple gaussian mode profile ($l = 0$) or in the first-order LG modes ($l = \pm 1$) after astigmatic mode conversion (for a description of this technique, see ref. 22). Spontaneous parametric down-conversion was done in a 1.5-mm-thick BBO (β -barium borate) crystal cut for type-I phase matching (that is, both photons carry the same linear polarization). The crystal cut was chosen so as to produce down-converted photons at a wavelength of 702 nm at an angle of 4° off the pump direction.

The mode detection of the down-converted photons was performed for gaussian and LG modes. The gaussian mode ($l = 0$) was identified using mono-mode fibres (Fig. 2) in connection with avalanche detectors. All other modes have a larger spatial extension, and therefore cannot be coupled into the mono-mode fibre. The LG modes ($l \neq 0$) were identified using mode detectors consisting of computer-generated holograms and mono-mode optical fibres (Fig. 2).

Computer-generated holograms have often been exploited for creating LG modes of various orders²³. Our holograms were phase gratings $5 \text{ mm} \times 5 \text{ mm}$ in size, with 20 lines mm^{-1} , which we first recorded on holographic films and bleached afterwards to increase the transmission efficiency (Fig. 2). We made holograms that had one or two dislocations in the centre, and designed them to have their maximum intensity in the first diffraction order. This enabled us to distinguish between LG modes $l = -2, -1, 0, 1, 2$ using all holograms in the first diffraction order—for which they have been blazed. For analysing an LG mode with a negative index, the holograms were rotated by 180° around the axis perpendicular to the grating lines. The total transmission efficiency of all our holograms was about 80%, and they diffracted 18% of the incoming beam into the desired first order. These characteristics were measured at 632 nm wavelength because a laser source at 702 nm was not available.

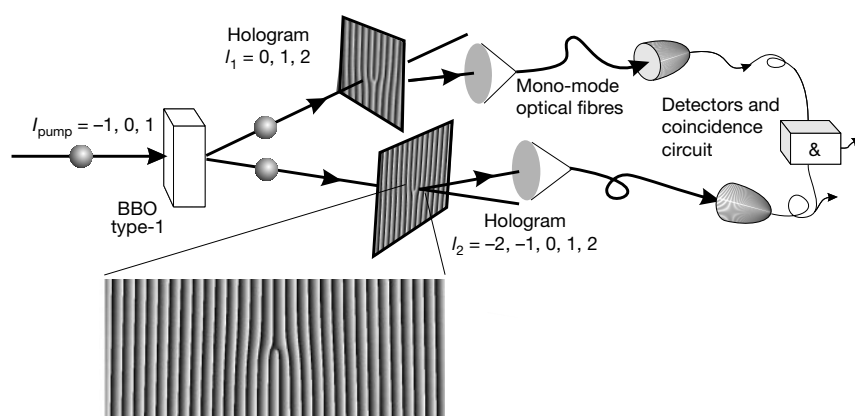


Figure 2 Experimental set-up for single-photon mode detection. After parametric down-conversion, each of the photons enters a mode detector consisting of a computer-generated hologram and a mono-mode optical fibre. By diffraction at the hologram, the incoming mode undergoes a mode transformation in such a way that an LG mode can be transformed into a gaussian mode. As it has a smaller spatial extension than all LG modes, only the gaussian mode can be coupled into the mono-mode fibre. Thus observation of a click projects the mode incident on the fibre coupler into the gaussian mode. The hologram is a phase grating with Δm dislocations in the centre blazed for first-order diffraction. An incoming gaussian laser beam passing through the dislocation of the hologram is diffracted by the grating, and the n th diffraction order becomes an LG mode

The diffraction efficiency is not the only loss that occurs. We also have to account for Fresnel losses at all optical surfaces (95% transmission), imperfect coupling into the optical fibres (70% for a gaussian beam), non-ideal interference filters (75% centre transmission), and the efficiency of the detectors (30%). A conservative estimate of all the losses yields an overall collection efficiency of 2–3%. Comparing the unnormalized ($l_{\text{pump}} = l_1 = l_2 = 0$) coincidence rates of about $2,000 \text{ s}^{-1}$ to the single count rates of about $100,000 \text{ s}^{-1}$ we deduce an efficiency of 2%, in agreement with the above estimation.

The mode analysis was performed in coincidence for all cases where mode filter 1 was prepared for analysing LG modes $l_1 = 0, 1, 2$ and mode filter 2 for those with $l_2 = -2, -1, 0, 1, 2$. For analysing an LG mode with mode index $l = 0$ —that is, a gaussian mode—the dislocation of the hologram was shifted out of the beam path. Thus the beam was sent through the border of the hologram where it acts as a conventional grating without changing the photons' angular momentum. The results are shown in Fig. 3 for different values of orbital angular momenta of the pump beam. Within experimental accuracy, coincidences were only observed in those cases where the sum of the orbital angular momenta of the down-converted photons was equal to the pump beam's orbital angular momentum. However, the absolute count rates of these cases are not equal. This is probably due to unequal emission probabilities of the photons into the different modes in the down-conversion process.

These results confirm conservation of the orbital angular momentum in parametric down-conversion. The signal-to-noise ratios achieved were as high as $V = 0.976 \pm 0.038$ and $V = 0.916 \pm 0.009$ for pump beams with and without orbital angular momentum, respectively. V is defined as $(I_{\text{out}} - I_{\text{in}})/(I_{\text{out}} + I_{\text{in}})$, where I_{in} and I_{out} denote the maximum and the minimum of the coincidences with the dislocation of the hologram respectively in and out of the beam.

It is only by using a coincidence measurement that we could show that the conservation of the orbital angular momentum holds for each single photon pair. In contrast, cumulative detection methods using many photons result in an incoherent pattern²⁴, because

with an index $l = n\Delta m$ and vice versa. Intuitively speaking, the phase dislocation exerts a 'torque' onto the diffracted beam because of the difference of the local grating vectors in the upper and lower parts of the grating. This 'torque' depends on the diffraction order n and on Δm . Consequently the right and left diffraction orders gain different handedness. Reversing this process, a photon with angular momentum $\Delta m/\hbar$ before the grating can be detected by the mono-mode fibre detector placed in the first diffraction order. A photon with zero angular momentum (gaussian mode) is detected by diffracting the beam at the border of the hologram far away from the dislocation. All our measurements were performed in coincidence detection between the two down-converted photons.

each beam from parametric down-conversion by itself is an incoherent mixture. Therefore some previous workers²⁴ using these classical detection methods—which are in principle unsuitable at the single photon level—were led to believe that the orbital angular momentum is not conserved in spontaneous parametric down-conversion.

Given this experimental verification of the conservation of orbital angular momentum, entanglement between the two photons produced in the conversion process might be expected. But to explain the conservation of the orbital angular momentum, the photons do not necessarily have to be entangled: it would be sufficient to assume classical correlation. But further experimental results (see below) showed that the two-photon state goes beyond classical correlation, and indeed, we were able to prove the entanglement for photon states with phase singularities.

To confirm entanglement, we have to demonstrate that the two-photon state is not just a mixture but a coherent superposition of product states of the various gaussian and LG modes which obey angular momentum conservation. For simplicity, we restricted ourselves to superpositions of two basis states only. An important distinction between coherent superposition and incoherent mixture of gaussian and LG modes is that the latter possess no phase singularity. This is because adding the spatial intensity distributions of these two modes will yield a finite intensity everywhere in the resulting pattern. In contrast, in a coherent superposition the amplitudes are added, and therefore the phase singularity must remain and is displaced to an eccentric location (Fig. 4). It will appear at that location where the amplitudes of the two modes are equal, with opposite phase. Therefore the radial distance of the singularity from the beam centre is a measure of the amplitude ratio of the gaussian to the LG components, whereas the angular position of the singularity is determined by their relative phase. Intuitively speaking, the position of the dislocation with respect to the beam is equivalent to the orientation of a polarizer.

Superpositions of LG and gaussian modes can be realized experimentally by shifting the dislocation of the hologram out of the centre of the beam by a certain (small) amount. Hence in order to detect a photon having an orbital angular momentum that is a superposition of the gaussian and the LG mode, the hologram was placed in a position such that the dislocation was slightly displaced from the beam centre. In the intensity pattern these modes possess an eccentric singularity (Fig. 4). To demonstrate the entanglement, we therefore shifted one of the holograms and scanned the gaussian mode filter on the other side while recording the coincidences.

The results shown in Fig. 4 verify the correlation in superposition bases of the LG ($l = \pm 2$) and gaussian ($l = 0$) modes. A closer analysis shows that there are two conditions necessary to obtain the

measured curves. First, the shifted hologram has to work as described above, and second, the source must emit an angular-momentum-entangled state. Assume that the source only emits classically correlated but not entangled singularities. Then on the side with the shifted hologram, the various terms of the classical mixture would be projected onto a state with displaced singularity leaving the total state again in a mixture. Respecting the conservation of angular momentum we would then have to sum the probabilities of the various components on the other side, resulting

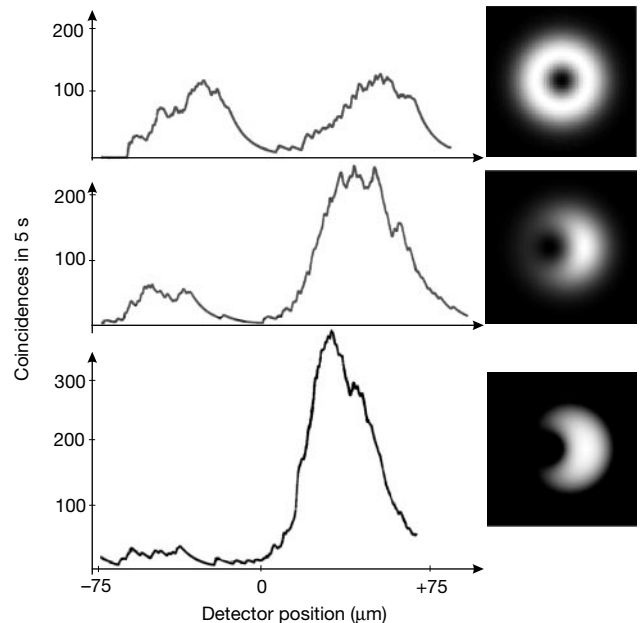


Figure 4 Experimental evidence (left; right, simulation) of entanglement of photon states with phase singularities. The dislocation of the hologram in the beam of photon 1 is shifted out of the beam centre step by step (top, middle, bottom). In these positions, this hologram—together with the mono-mode fibre detector—projects the state of photon 1 into a coherent superposition of LG and gaussian modes. The mode filter for photon 2 with the hologram taken out makes a scan of the second photon's intensity distribution (detector position) in order to identify the location of its singularity with respect to the beam centre. The coincidences show that the second photon is also detected in a superposition of LG and gaussian modes. Classical correlation would yield a coincidence picture which is just a mixture of gaussian and LG modes. In that case, the intensity minimum would remain in the beam centre but would become washed out. In the experiment a hologram with two dislocations in the first diffraction order was used.

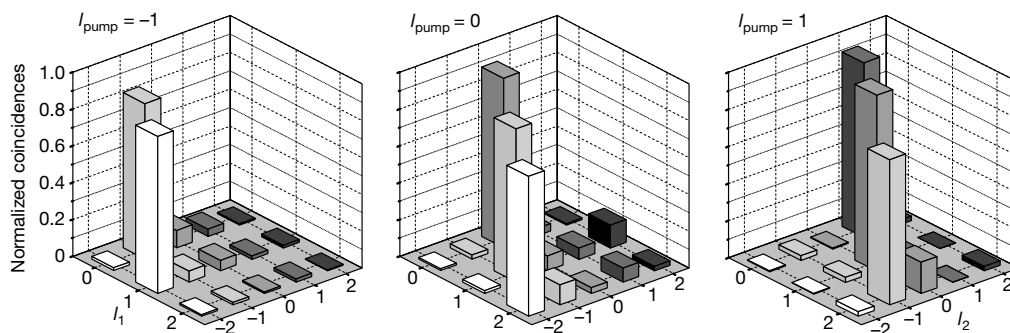


Figure 3 Conservation of orbital angular momentum. Coincidence mode detections for photon 1 and photon 2 in 15 possible combinations of orthogonal states were performed. This was done for a pump beam having an orbital angular momentum of $-\hbar$, 0 and $+\hbar$ per photon, respectively. Coincidences were observed in all cases where the sum of the

orbital angular momenta of the down-converted photons was equal to the pump beam's orbital angular momentum. The coincidence counts for each fixed value of the orbital angular momentum of photon 1 were normalized by the total number of coincidences varying the orbital angular momentum of photon 2.

in a coincidence pattern not containing any intensity zeroes. Such a coincidence pattern would also be observed if a shifted hologram together with a mono-mode detector were not able to analyse for superposition states.

An entangled state represents correctly both the correlation of the eigenmodes and the correlations of their superpositions. Having experimentally confirmed the quantum superposition for $l = 0$ and $l = \pm 2$, it is reasonable to expect the quantum superposition will also occur for the other states. Nevertheless, ultimate confirmation of entanglement will be a Bell inequality experiment generalized to more states²⁵. Such an experiment will be a major experimental challenge, and we are preparing to perform it.

For a pump beam with zero angular momentum, the emitted state must then be represented by

$$\psi = C_{0,0}|0\rangle|0\rangle + C_{1,-1}|1\rangle - 1\rangle + C_{-1,1}| - 1\rangle|1\rangle + C_{2,-2}|2\rangle - 2\rangle + C_{-2,2}| - 2\rangle|2\rangle + \dots \quad (1)$$

as the LG modes form an infinite dimensional basis. Here the numbers in the brackets represent the indices l of the LG modes, and the C_{ij} denote the corresponding probability amplitude for measuring $|i\rangle|j\rangle$. The state (1) is a multi-dimensional entangled state for two photons, which in general will also contain terms with radial mode index $p \neq 0$. It means neither photon in state (1) possesses a well-defined orbital angular momentum after parametric down-conversion. The measurement of one photon defines its orbital angular momentum state, and projects the second one into the corresponding orbital angular momentum state.

It is conceivable that these states could in the future be extended to multi-dimensional multi-particle entanglement. A growing body of theoretical work calls for entanglement of quantum systems of higher dimensions^{7,8}. These states have applications in quantum cryptography with higher alphabets and in quantum teleportation. As such states increase the flux of information, it is conceivable that they could be important for many other applications in quantum communication and in quantum information. The possibility of using these photon states to drive micromachines, and the application of these states as optical tweezers, make them versatile and potentially suitable for future technologies^{19–21}. □

Received 12 March; accepted 5 June 2001.

1. Schrödinger, E. Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften* **23**, 807–812; 823–828; 844–849 (1935).
2. Schrödinger, E. Discussion of probability relations between separated systems. *Proc. Camb. Phil. Soc.* **31**, 555–563 (1935).
3. Bouwmeester, D., Pan, J.-W., Daniell, M., Weinfurter, H. & Zeilinger, A. Observation of a three-photon Greenberger-Horne-Zeilinger state. *Phys. Rev. Lett.* **82**, 1345–1349 (1999).
4. Pan, J.-W., Bouwmeester, D., Daniell, M., Weinfurter, H. & Zeilinger, A. Experimental test of quantum nonlocality in three-photon Greenberger-Horne-Zeilinger entanglement. *Nature* **403**, 515–519 (2000).
5. Sackett, C. A. *et al.* Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
6. Pan, J.-W., Daniell, M., Gasparoni, S., Weihs, G. & Zeilinger, A. Experimental demonstration of four-photon entanglement and high-fidelity teleportation. *Phys. Rev. Lett.* **86**, 4435–4438 (2001).
7. DiVincenzo, D. P., More, T., Shor, P. W., Smolin, J. A. & Terhal, B. M. Unextendible product bases, uncompletable product bases and bound entanglement. Preprint quant-ph/9908070 at <http://xxx.lanl.gov> (1999).
8. Bartlett, S. D., de Guise, H. & Sanders, B. C. Quantum computation with harmonic oscillators. Preprint quant-ph/0011080 at <http://xxx.lanl.gov> (2000).
9. Bechmann-Pasquinucci, H. & Peres, A. Quantum cryptography with 3-state systems. *Phys. Rev. Lett.* **85**, 3313–3316 (2000).
10. Bechmann-Pasquinucci, H. & Tittel, W. Quantum cryptography using larger alphabets. *Phys. Rev. A* **61**, 62308–62313 (2000).
11. Bourennane, M., Karlsson, A. & Björk, G. Quantum key distribution using multilevel encoding. *Phys. Rev. A* (in the press).
12. Reck, M., Zeilinger, A., Bernstein, H. J. & Bertani, P. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.* **73**, 58–61 (1994).
13. Zukowski, M., Zeilinger, A. & Horne, M. Realizable higher-dimensional two-particle entanglements via multiport beam splitters. *Phys. Rev. A* **55**, 2564–2579 (1997).
14. Reck, M. *Quantum Interferometry with Multiports: Entangled Photons in Optical Fibers*. Thesis, Univ. Innsbruck (1996).
15. Arnaut, H. H. & Barbosa, G. A. Orbital and angular momentum of single photons and entangled pairs of photons generated by parametric down-conversion. *Phys. Rev. Lett.* **85**, 286–289 (2000).

16. Franke-Arnold, S., Barnett, S. M., Padgett, M. J. & Allen, L. Two-photon entanglement of orbital angular momentum states. *Phys. Rev. A* (in the press).
17. Allen, L., Beijersbergen, M. W., Spreeuw, R. J. C. & Woerdman, J. P. Orbital angular momentum of light and the transformation of laguerre-gaussian laser modes. *Phys. Rev. A* **45**, 8185–8189 (1992).
18. He, H., Fries, M., Heckenberg, N. & Rubinsztein-Dunlop, H. Direct observation of transfer of angular momentum to absorptive particles from a laser beam with a phase singularity. *Phys. Rev. Lett.* **75**, 826–829 (1995).
19. Simpson, N. B., Dholakia, K., Allen, L. & Padgett, M. J. Mechanical equivalence of spin and orbital angular momentum of light: An optical spanner. *Opt. Lett.* **22**, 52–54 (1997).
20. Galajda, P. & Ormos, P. Complex micromachines produced and driven by light. *Appl. Phys. Lett.* **78**, 249–251 (2001).
21. Friese, M. E. J., Enger, J., Rubinsztein-Dunlop, H. & Heckenberg, N. Optical angular-momentum transfer to trapped absorbing particles. *Phys. Rev. A* **54**, 1593–1596 (1996).
22. Beijersbergen, M. W., Allen, L., van der Veen, H. E. L. O. & Woerdman, J. P. Astigmatic laser mode converters and transfer of orbital angular momentum. *Opt. Commun.* **96**, 123–132 (1993).
23. Arit, J., Dholakia, K., Allen, L. & Padgett, M. J. The production of multiringed laguerre-gaussian modes by computer-generated holograms. *J. Mod. Opt.* **45**, 1231–1237 (1998).
24. Arit, J., Dholakia, K., Allen, L. & Padgett, M. Parametric down-conversion for light beams possessing orbital angular momentum. *Phys. Rev. A* **59**, 3950–3952 (1999).
25. Kaszlikowski, D., Gnacinski, P., Zukowski, M., Miklaszewski, W. & Zeilinger, A. Violation of local realism by two entangled n -dimensional systems are stronger than for two qubits. *Phys. Rev. Lett.* **85**, 4418–4421 (2000).

Acknowledgements

This work was supported by the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (FWF).

Correspondence and requests for materials should be addressed to A.Z. (e-mail: anton.zeilinger@univie.ac.at).

Superconductivity in the non-magnetic state of iron under pressure

Katsuya Shimizu*†, Tomohiro Kimura*, Shigeyuki Furomoto*, Keiki Takeda*, Kazuyoshi Kontani*, Yoshichika Onuki‡ & Kiichi Amaya*†

* Department of Physical Science, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, USA

† Research Center for Materials Science at Extreme Conditions, Osaka University, Toyonaka, Osaka 560-8531, Japan

‡ Department of Physics, Graduate School of Science, Osaka University, Osaka 560-0043, Japan

Ferromagnetism and superconductivity are thought to compete in conventional superconductors, although in principle it is possible for any metal to become a superconductor in its non-magnetic state at a sufficiently low temperature. At pressures above 10 GPa, iron is known to transform to a non-magnetic structure^{1,2} and the possibility of superconductivity in this state has been predicted^{3,4}. Here we report that iron does indeed become superconducting at temperatures below 2 K at pressures between 15 and 30 GPa. The transition to the superconducting state is confirmed by both a drop in resistivity and observation of the Meissner effect.

An iron sample with purity of 99.995% (Johnson Matthey) was purified further and degassed by heating close to the melting point in an ultra-high-vacuum chamber. The sample was cut into a rectangular shape of $0.04 \times 0.160 \times 0.07 \text{ mm}^3$ and placed in the sample chamber of a non-magnetic diamond-anvil cell (DAC) made of BeCu alloy. For electrical resistivity measurements the BeCu metal gasket was covered with a thin Al_2O_3 layer for electrical insulation. Electrical resistivity measurements are performed using the a.c. four-terminal method with a typical measuring current of $0.1 \times 10^{-6} \text{ A}$ at low temperatures below 10 K. The sample chamber was filled with NaCl as the pressure-transmitting medium (Fig. 1a). Several ruby chips of less than 0.002 mm in diameter were located around the sample and the applied pressure

was determined by a standard ruby fluorescence method. The pressure was applied and fixed at room temperature, and the DAC was assembled on the mixer of a $^3\text{He}/^4\text{He}$ dilution refrigerator; the temperature dependence of the resistivity of iron was measured at temperatures down to 30 mK.

A small but definite drop in the electrical resistivity was observed at temperatures below 2 K at pressures above 16 GPa. A typical resistivity–temperature (R – T) curve at a pressure of 25 GPa is shown in Fig. 1b. As expected for conventional superconductors, the drop was suppressed and broadened by a large measuring current above 1×10^{-6} A. When we applied an external magnetic field the resistivity drop shifted towards lower temperatures and disappeared at 0.2 T, as shown in Fig. 2. The observed magnetic-field dependence shows that the drop was due to the superconducting transition of the iron sample. The measured critical field for iron is almost ten times larger than that of other superconductors such as rhenium or thallium with similar critical temperatures of about 2 K. The change in the resistivity due to the transition was less than 10%

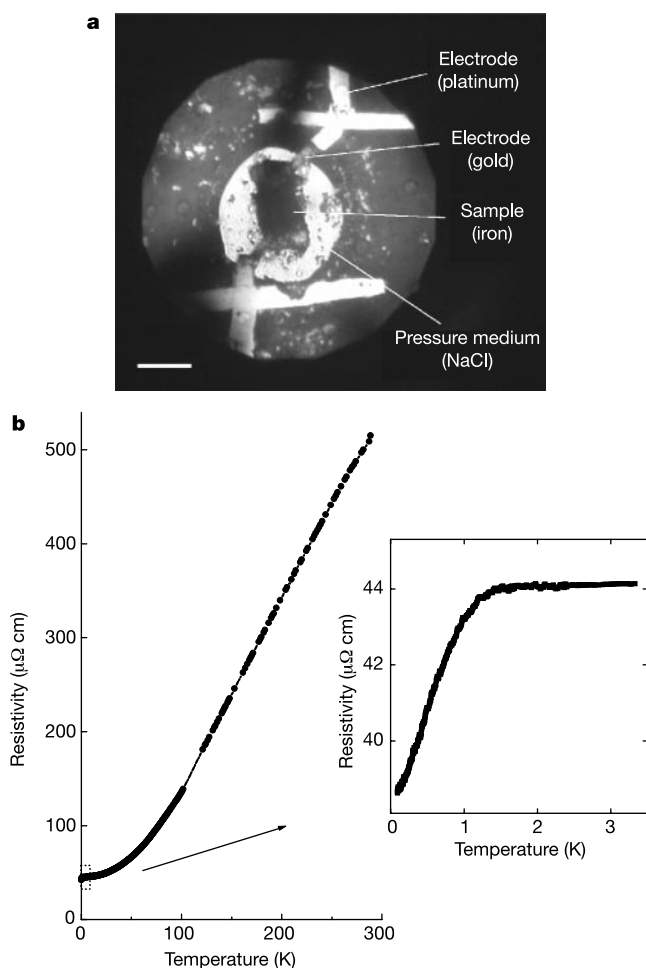


Figure 1 Arrangement of sample in the pressure-cell, and the low-temperature experimental curve of the resistance of iron at 25 GPa. **a**, Photograph showing the configuration of the sample and electrodes on the pressure surface of the diamond anvil. The gold wires, 0.01 mm in diameter, were attached to the sample crystal using micro-spot welding and placed in the sample chamber, which was filled with sodium chloride and connected to a platinum electrode outside the chamber. **b**, Temperature dependence of the electrical resistivity of iron at 25 GPa. A 10% drop in resistivity indicates the onset of superconductivity at around 1.5 K. To our knowledge, no report exists for the superconductivity of gold or platinum, including the pressure-induced case. Nor had we previously detected a superconducting signal from gold or platinum in experiments in this pressure region. Scale bar, 0.1 mm.

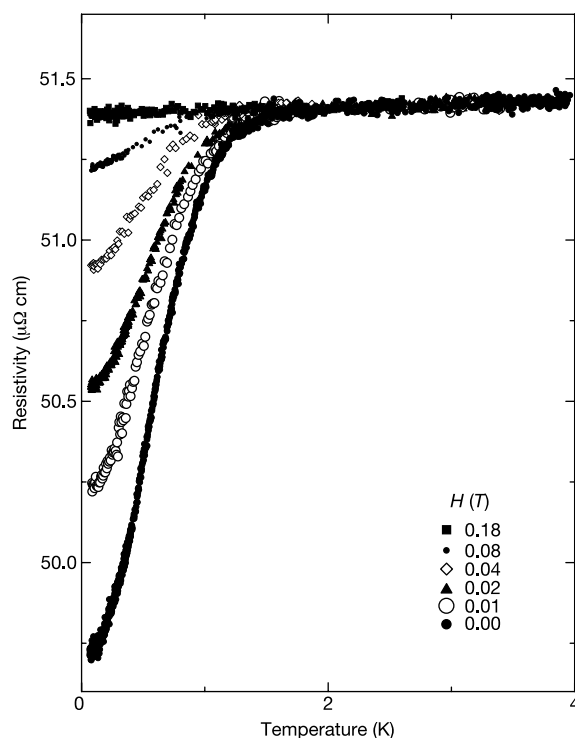


Figure 2 Magnetic-field dependence of the resistance drop of iron for various magnetic fields at 23 GPa. The resistance drop shifts towards lower temperature with increasing applied magnetic field. Above 0.18 T, only the temperature-independent residual resistance value is observed.

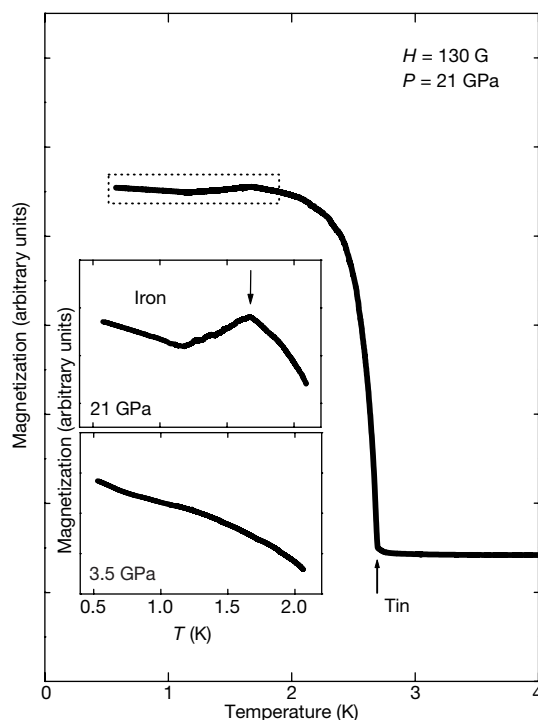


Figure 3 The temperature dependence of the magnetization of iron under pressure obtained by cooling the sample at a magnetic field of 130 G. The signal at 21 GPa (the area enclosed by the dotted line is expanded in the upper inset) shows the appearance of diamagnetism at temperatures below 1.7 K, which is confirmed by the signal given by tin at 2.7 K. The lower inset shows the disappearance of the Meissner signal in iron when the pressure is decreased to 3.5 GPa in the b.c.c. phase.

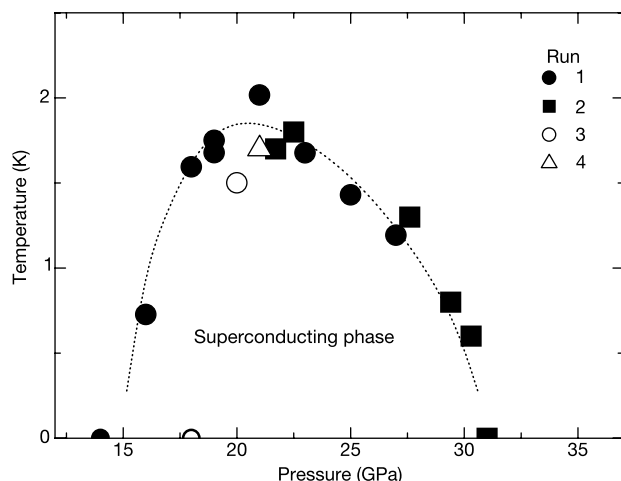


Figure 4 Superconducting phase diagram of iron. The transition temperature, T_c , is defined by the onset of the transition. The dashed curve is a guide for the eye. T_c values shown by black circles and squares (runs 1 and 2) are obtained from electrical resistance measurements with sodium chloride as a pressure medium. Run 3 (white circle) is a run without a pressure medium while releasing the pressure from 90 GPa. Run 4 (white triangle) shows the onset of the Meissner signal in the magnetization measurement. T_c appears at around 15 GPa and increases with pressure. T_c reaches the maximum value of 2 K at around 21 GPa and vanishes above 30 GPa.

of the total residual resistivity. This is because the residual resistivity includes contributions from the thin gold wire in series with the sample as well as from the finite contact resistance, which is estimated to be around 90% of the measured resistance. The main impurities contained in the starting material were O, Ta, Si and Co at concentrations of 115, 10, 8.3 and 8.3 p.p.m., respectively, according to the chemical analysis data. Thus we regard the drop as the onset of superconductivity of iron.

To confirm the superconductivity, we tried to detect the Meissner effect using a SQUID (superconducting quantum interference device) magnetometer at 21 GPa. A 4:1 mixture of methanol/ethanol was used as the pressure-transmitting medium. An iron sample was obtained from the rod that provided the samples used for the resistance measurements, cut into slices 0.04 mm in thickness and 0.15 mm in diameter, and placed in the chamber of the BeCu gasket. Several turns of pick-up coil for the SQUID magnetometer were wound closely around the pressure surface of the diamond and an equal number of turns near to the diamond for background compensation. A small tin chip was put inside the compensation coil as a reference sample to check both the sign and the magnitude of the signal from iron. Figure 3 shows the superconducting transition signals from iron at 1.7 K and tin at 2.7 K, of opposite sign.

The pressure dependence of the transition temperature T_c was observed in the pressure range between 15 GPa and 30 GPa (Fig. 4). Superconductivity was observed in three different runs of the resistivity measurements and the magnetization measurement. Runs 1 and 2 were obtained by quasi-hydrostatic conditions with NaCl as the pressure-medium. No pressure-medium was used for run 3 and the superconductivity transition was observed only while reducing the pressure from 90 GPa. This implies that even a very small amount of remaining ferromagnetic b.c.c. iron may suppress the onset of superconductivity and that its transformation to h.c.p. iron is caused by either quasi-hydrostatic pressure or by increasing the pressure as much as possible. We conclude that the superconducting phase of iron appears after the establishment of the non-magnetic state at 15 GPa, which is close to the beginning of the b.c.c.-to-h.c.p. structural transition of iron. □

Received 29 January; accepted 25 May 2001.

1. Cort, G., Taylor, R. D. & Willis, J. O. Search for magnetism in hcp ϵ -Fe. *J. Appl. Phys.* **53**, 2064–2065 (1982).
2. Taylor, R., Pasternak, M. & Jeanloz, R. Hysteresis in the high pressure transformation of bcc- to hcp-iron. *J. Appl. Phys.* **69**, 6126–6128 (1991).
3. Wohlfarth, E. P. The possibility that ϵ -Fe is a low temperature superconductor. *Phys. Lett. A* **75**, 141–143 (1979).
4. Freeman, A. J., Continenza, A., Massida, S. & Grossman, J. C. Structural and electronic properties and possible superconductivity in simple cubic Fe. *Physica C* **166**, 317–322 (1990).

Acknowledgements

We thank T. C. Kobayashi for stimulating discussion. This work has been supported by a Grant-in-Aid for COE (Center of Excellence) Research and Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology in Japan, and also by Core Research for Evolution Science and Technology of the Japan Science and Technology Corporation.

Correspondence and requests for materials should be addressed to K.S. (e-mail: kshimizu@mp.es.osaka-u.ac.jp).

Direct observation of hole transfer through DNA by hopping between adenine bases and by tunnelling

Bernd Giese, Jérôme Amaudrut, Anne-Kathrin Köhler, Martin Spormann & Stephan Wessely

Department of Chemistry, University of Basel, St Johannis Ring 19, CH-4056 Basel, Switzerland

The function of DNA during oxidative stress¹ and its suitability as a potential building block for molecular devices^{2–4} depend on long-distance transfer of electrons and holes through the molecule, yet many conflicting measurements of the efficiency of this process have been reported^{5,6}. It is accepted that charges are transported over long distances through a multistep hopping reaction^{7–11}; this ‘G-hopping’⁸ involves positive charges moving between guanines (Gs), the DNA bases with the lowest ionization potential. But the mechanism fails to explain the persistence of efficient charge transfer when the guanine sites are distant^{7,12}, where transfer rates do not, as expected, decrease rapidly with transfer distance. Here we show experimentally that the rate of charge transfer between two guanine bases decreases with increasing separation only if the guanines are separated by no more than three base pairs; if more bridging base pairs are present, the transfer rates exhibit only a weak distance dependence. We attribute this distinct change in the distance dependence of the rate of charge transfer through DNA to a shift from coherent superexchange charge transfer (tunnelling) at short distances to a process mediated by thermally induced hopping of charges between adenine bases (A-hopping) at long distances. Our results confirm theoretical predictions of this behaviour^{13–17}, emphasizing that seemingly contradictory observations of a strong^{8,9} as well as a weak^{7,12} influence of distance on DNA charge transfer are readily explained by a change in the transfer mechanism.

We have measured the efficiency of the charge transfer between Gs, separated by adenine-thymine (A-T)_n bridges of various lengths, in double strands **1a–h** (Fig. 1). Photolysis of the 4'-acylated nucleotide in DNA **1** generates the sugar radical cation in **2** (Fig. 1), which injects a positive charge into G₂₂ of the complementary, radiolabelled strand (**2**→**3**; Fig. 1). This guanine radical cation G₂₂⁺ is either trapped irreversibly by water, yielding after piperidine treatment the strand cleavage product P_G, or it induces an electron

(hole) transfer through DNA (3→4; Fig. 1). In order to build up a driving force for the charge transfer process, we used as electron donor a GGG triplet that is more easily oxidized than a single G¹⁸. The positive charge reaching the GGG unit was quantified by the

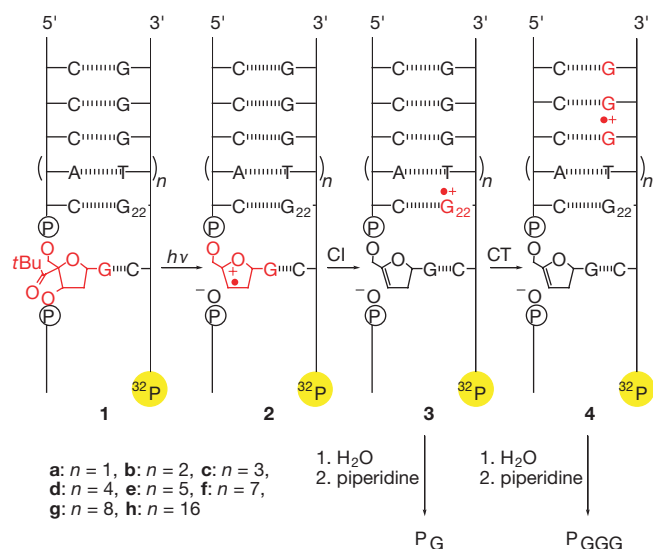


Figure 1 Charge injection and charge transfer in DNA strands. Illustrated are the method of charge injection (CI) into G₂₂ (2→3) and charge transfer (CT) to the GGG sequence (3→4), starting from DNA strands **1a–h** that contain a 4'-acylated deoxyguanoside (in red). The double strands above and below the depicted sequences are identical to those used in earlier studies¹⁹. Photolyses of **1a–h** were performed at pH 5.0 (citrate buffer) with a 500-W, high-pressure Hg lamp (320 nm cut-off filter) in the absence of oxygen at 15 °C. Trapping by water of the charge at G₂₂ and GGG leads to cleavage products P_G and P_{GGG}, respectively, after piperidine treatment. The nucleotides in red indicate the charge precursor (**1**), the injection site (**2**), the charge donor (**3**), and the charge acceptor (**4**).

water trapping product P_{GGG}, and the yield ratios P_{GGG}/P_G were determined by gel electrophoresis. We observed the product ratios shown in Fig. 2; to obtain these results, we used the assay of ref. 19 but changed the pH from 7.0 to 5.0.

The data in Fig. 2 show that the efficiency of the charge transfer, measured by the P_{GGG}/P_G ratio, drops by a factor of 8 for each additional A·T base pair in short (A·T)_n bridges (n = 1–3). But in longer sequences (n = 4–7), the P_{GGG}/P_G ratio decreases only very slightly when the number of A·T base pairs increases (Fig. 2). At further elongation of the (A·T)_n sequence, the distance influence vanishes completely. Thus, we observed no change in the P_{GGG}/P_G ratio by increasing the number of A·T base pairs from n = 7 to n = 16. A plot of log P_{GGG}/P_G against the number n of the A·T base pairs between the guanines exhibits a clear switch of the influence of distance on the charge transfer (Fig. 3).

Such a distance dependence, which demonstrates a change of the reaction mechanism, has been predicted recently^{13–17}. From being a coherent superexchange charge transfer (tunnelling) process at short distances, the mechanism becomes a thermally induced hopping process for long (A·T)_n sequences, where the adenines are involved as charge carriers (A-hopping)¹⁷. A switch between these reaction mechanisms occurs because tunnelling rates decrease considerably as the distance increases (see the steep line at low n in Fig. 3)^{8,9}. Therefore, in DNA strands where the guanines are separated from each other by long (A·T)_n sequences, endothermic²⁰ transfer of the positive charge from a guanine radical cation (G⁺) to an adjacent adenine becomes faster than the direct transfer of this charge to the distant guanine. The subsequent migration of the positive charge between the adenines (A-hopping) is so rapid that the length of the (A·T)_n sequence plays only a minor role. Further experiments with DNA double strands **5** and **6** show that, in contrast to the behaviour seen with guanine radical cations, trapping of adenine radical cations by water is insignificant (Fig. 4). This difference is expected¹⁷, because the water trapping reaction of adenine radical cations proceeds through a transition state, which is much higher in energy than the transition state associated with

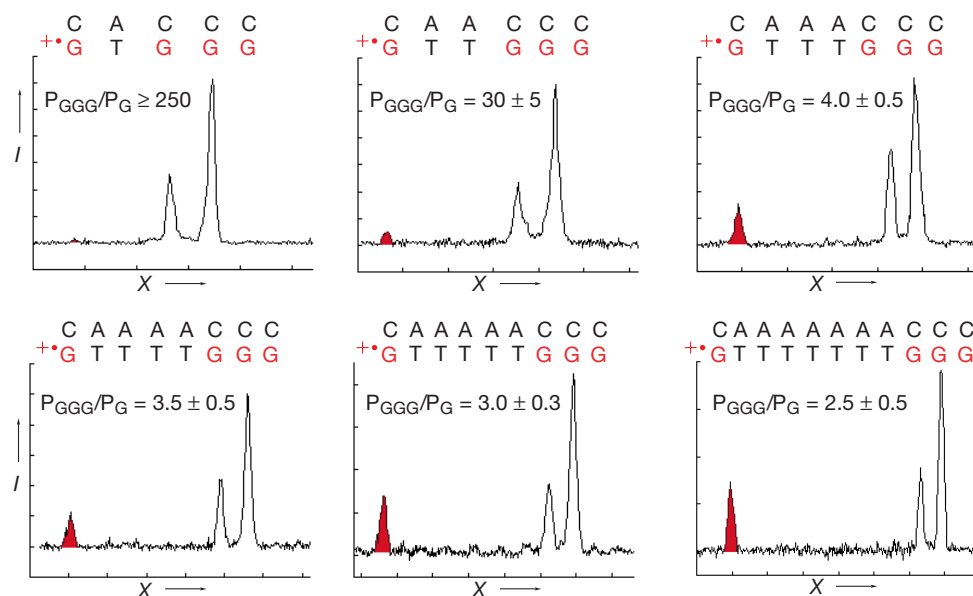


Figure 2 Ratios of the irradiation products derived from initially formed guanine radical cations to those derived after charge transfer across (A·T)_n bridges. The individual curves show radioactivity intensities of radiolabelled DNA strands subjected to gel electrophoresis using denaturing polyacrylamide gels. They are obtained by subtraction of intensities measured in control experiments (irradiation of unmodified strands) from intensities measured in irradiation experiments with the modified strands **1a–h**. The

peaks in intensity (I) correspond to products P_G (shown in red) and P_{GGG}. The peak areas are relative yields, thus the P_{GGG}/P_G ratios are yield ratios. The products are formed by water trapping of the guanine radical cations G⁺ and subsequent site-selective strand cleavage of the radiolabelled strand. Each peak arises at the nucleotide position (X) shown in the strand depicted at the top of each panel. The nucleotides in red indicate the beginning and the end of the charge transfer process.

the trapping reaction of the guanine radical cation. The data of Fig. 4 also confirm¹⁷ that interstrand A-hopping steps lead to a slight decrease of the hole transfer efficiency.

Our experiments show the presence of two different processes for the hole transfer between guanines in DNA: (1) a coherent superexchange reaction (single-step tunnelling), where the bridging

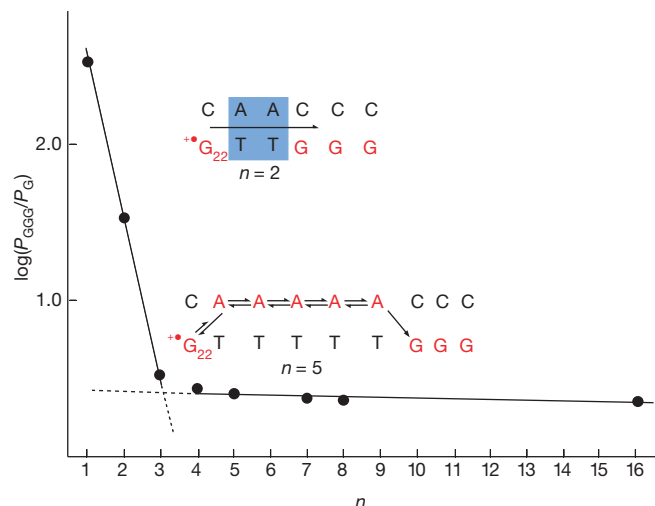


Figure 3 Plot of $\log(P_{\text{GGG}}/P_{\text{G}})$ against the number n of the A-T base pairs. Each experiment was performed three times, and their relative errors are within ± 10 –20% (see Fig. 2). The steep line corresponds to the coherent superexchange charge transfer. Its slope leads to $(\beta = 0.6 \text{ \AA}^{-1})$ of the Marcus-Levich-Jortner equation¹⁰. The flat line is drawn in order to make clear the weak distance dependence. The product ratios $P_{\text{GGG}}/P_{\text{G}}$ are proportional to the charge transfer rates. The arrows in the depicted DNA strands indicate the superexchange charge transfer between G_{22}^{+} and the GGG sequence for short distances ($n = 2$), or the A-hopping mechanism for long distances ($n = 5$), where—in addition—adenines act as charge carriers. For clarity, only the double strands with $n = 2$ and $n = 5$ are shown. The nucleotides in red indicate all charge carriers.

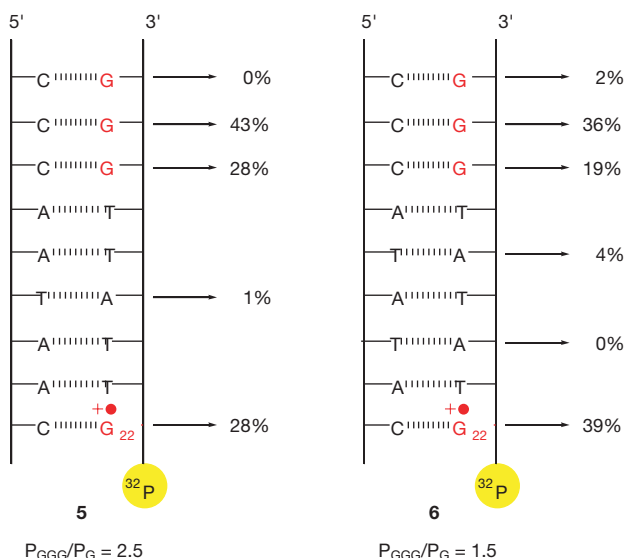


Figure 4 Relative yields of water trapping products during hole transfer through DNA sequences **5** and **6**. The experiments were performed three times (strand **5**) or twice (strand **6**). The relative errors of yields of products P_{GGG} and P_{G} are $\pm 15\%$. The yields obtained by trapping the A sites have absolute errors of $\pm 4\%$. Trapping products of thymine bases could not be detected. The nucleotides in red indicate the beginning and the end of the charge transfer process.

adenines are indirectly affecting the transfer mechanism by mediating the electronic coupling between the guanines, and (2) a thermally induced hopping process (A-hopping mechanism), where the lifetime of the guanine radical cation is long enough to oxidize the intervening adenine bases and directly involve them in charge transport. The efficiency of the tunnelling reaction decreases rapidly with the number of the intervening A-T base pairs, whereas the A-hopping process is only slightly influenced by the number of the A-T base pairs. Thus, the debated contradiction between very strong^{8,9} and very weak^{7,12} influence of distance on the hole transfer through DNA can be explained by a change in the reaction mechanism.

These results indicate that A-hopping provides a mechanism for ameliorating the harmfulness of damage to DNA under conditions of oxidative stress, by providing a means for charge transport over long sequences of A-T pairs into regions of low ionization potentials^{1,21}. In addition, the ability to rationally change the behaviour of a system from strongly distance-dependent charge transfer to weakly distance-dependent charge transfer may open new opportunities in nanoelectronics. DNA can offer this ability because the switch to the activated hopping mechanism depends on the difference of the oxidation potentials of the DNA bases involved—and this difference could be easily changed by using suitable substituents. □

Received 19 March; accepted 6 June 2001.

- Heller, A. Spiers Memorial Lecture: On the hypothesis of cathodic protection of genes. *Faraday Discuss.* **116**, 1–13 (2000).
- Fink, H. W. & Schönenberger, C. Electrical conduction through DNA molecules. *Nature* **398**, 407–410 (1999).
- Porath, D., Bezryadin, A., De Vries, S. & Dekker, C. Direct measurement of electrical transport through DNA molecules. *Nature* **403**, 635–638 (2000).
- Kasumov, A. Y. *et al.* Proximity-induced superconductivity in DNA. *Science* **291**, 280–282 (2001).
- Lewis, F. D. *et al.* Distance dependent electron transfer in DNA hairpins. *Nature* **277**, 673–676 (1997).
- Murphy, C. J. *et al.* Long-range photoinduced electron transfer through a DNA helix. *Science* **262**, 1025–1029 (1993).
- Schuster, G. B. Long-range charge transfer in DNA: transient structural distortions control the distance dependence. *Acc. Chem. Res.* **33**, 253–260 (2000).
- Giese, B. Long-distance charge transport in DNA: the hopping mechanism. *Acc. Chem. Res.* **33**, 631–636 (2000).
- Lewis, F. D., Letsinger, R. L. & Wasielewski, M. R. Dynamics of photoinduced charge transfer and hole transport in synthetic DNA hairpins. *Acc. Chem. Res.* **34**, 159–170 (2001).
- Jortner, J., Bixon, M., Langenbacher, T. & Michel-Beyerle, M. E. Charge transfer and transport in DNA. *Proc. Natl Acad. Sci. USA* **95**, 12759–12765 (1998).
- Berlin, Y. A., Burin, A. L. & Ratner, M. A. Charge hopping in DNA. *J. Am. Chem. Soc.* **123**, 260–268 (2001).
- Williams, T. T., Odom, D. T. & Barton, J. K. Variations in DNA. Charge transport with nucleotide composition and sequences. *J. Am. Chem. Soc.* **122**, 9048–9049 (2000).
- Felts, A. K., Pollard, W. T. & Friesner, R. A. Multilevel Redfield treatment of bridge-mediated long-range electron transfer: a mechanism for anomalous distance dependence. *J. Phys. Chem.* **99**, 2929–2940 (1995).
- Okada, A., Chernyak, V. & Mukamel, S. Solvent reorganization in long-range electron transfer: density matrix approach. *J. Phys. Chem. A* **102**, 1241–1251 (1998).
- Segal, D., Nitzan, A., Davis, W. B., Wasielewski, M. R. & Ratner, M. A. Electron transfer rates in bridged molecular systems. A steady-state analysis of coherent tunneling and thermal transitions. *J. Phys. Chem. B* **104**, 3817–3829 (2000).
- Grozema, F. C., Berlin, Y. A. & Siebbeles, L. D. A. Mechanism of charge migration through DNA: molecular wire behavior, single-step tunneling or hopping? *J. Am. Chem. Soc.* **122**, 10903–10909 (2000).
- Giese, B. & Spichty, M. Long-distance charge transport through DNA: quantification and extension of the hopping model. *ChemPhysChem* **1**, 195–198 (2000).
- Saito, I. *et al.* Mapping of hot spots for DNA damage by one-electron oxidation: Efficacy of GG doublets and GGG triplets as a trap in long range hole migration. *J. Am. Chem. Soc.* **120**, 12686–12687 (1998).
- Giese, B. On the mechanism of long-range electron transfer through DNA. *Angew. Chem. Int. Edn Engl.* **38**, 996–998 (1999).
- Steenken, S. & Jovanovic, S. V. How easily oxidizable is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solution. *J. Am. Chem. Soc.* **119**, 617–618 (1997).
- Steenken, S., Jovanovic, S. V., Bietti, M. & Bernhard, K. The trap depth (in DNA) of 8-oxo-7,8-dihydro-2'-deoxyguanosine as derived from electron-transfer equilibria in aqueous solution. *J. Am. Chem. Soc.* **122**, 2373–2374 (2000).

Acknowledgements

We thank M. E. Michel-Beyerle and W. B. Davis for discussions. This work was supported by the Swiss National Science Foundation and the Volkswagen Foundation.

Correspondence and requests for materials should be addressed to B. G. (e-mail: bernd.giese@unibas.ch).

Thermolysis of fluoropolymers as a potential source of halogenated organic acids in the environment

David A. Ellis*, Scott A. Mabury*, Jonathan W. Martin† & Derek C. G. Muir‡

* Department of Chemistry, 80 St George Street, University of Toronto, Toronto, Ontario, Canada, M5S 3H6

† Department of Environmental Biology, University of Guelph, Guelph, Ontario, Canada, N1G 2W1

‡ National Water Research Institute, Environment Canada, 867 Lakeshore Road, Burlington, Ontario, Canada, L7R 4A6

Following the introduction of hydrochlorofluorocarbon (HCFCs) and hydrofluorocarbon (HFCs) gases as replacements for the ozone-destroying chlorofluorocarbons (CFCs), it has been discovered that HCFCs/HFCs can degrade in the atmosphere to produce trifluoroacetic acid¹, a compound with no known loss mechanisms in the environment^{2,3}, and higher concentrations in natural waters⁴ have been shown to be mildly phytotoxic⁵. Present environmental levels of trifluoroacetic acid are not accounted by HCFC/HFC degradation alone^{8–10}. Here we report that thermolysis of fluorinated polymers, such as the commercial polymers Teflon and Kel-F, can also produce trifluoroacetate and the similar compound chlorodifluoroacetate. This can occur either directly, or indirectly via products that are known to degrade to these haloacetates in the atmosphere¹¹. The environmental significance of these findings is confirmed by modelling, which indicates that the thermolysis of fluoropolymers in industrial and consumer high-temperature applications (ovens, non-stick cooking utensils and combustion engines) is likely to be a significant source of trifluoroacetate in urban rain water (~25 ng l⁻¹, as estimated for Toronto). Thermolysis also leads to longer chain polyfluoro- and/or polychlorofluoro- (C3–C14) carboxylic acids which may be equally persistent. Some of these products have recently been linked with possible adverse health⁶ and environmental impacts and are being phased out of the US market⁷. Furthermore, we detected CFCs and fluorocarbons—groups that can destroy ozone and act as greenhouse gases, respectively—among the other thermal degradation products, suggesting that continued use of fluoropolymers may also exacerbate stratospheric ozone-depletion and global warming.

The replacement CFC gases HCFC-123 (CF₃CHCl₂), HCFC-124 (CF₃CHClF) and HFC-134a (CF₃CH₂F) are known to degrade in the troposphere, through reaction with hydroxyl radicals, to trifluoroacetate, TFA¹. TFA is expected to be a long-lived environmental species, and has no known significant loss mechanism³. Studies have suggested that at higher concentrations, such as those observed in surface waters with seasonal evaporation–concentration cycles⁴, TFA will exhibit mild phytotoxicity⁵. Environmental modelling calculations conducted for TFA predict a concentration of 100–120 ng l⁻¹ in rain water between the years 2010 and 2020³; these predictions were based on the assumption that TFA originates from the CFC replacement gases. Environmental measurements^{8–10} of TFA have shown that current levels cannot be accounted for by these sources alone. In certain regions, concentrations of TFA in rainwater now average 120 ng l⁻¹ (ref. 8), and no major alternative source has been identified to explain these observations. Although there is some controversy within the literature, it seems that the TFA observed in precipitation is largely a result of urban activities. For example, TFA concentrations of <2–92 ng l⁻¹ have been found at remote sites, compared with 50–1,100 ng l⁻¹ for industrialized urban sites^{12,13}. The unidentified source of TFA is likely to be

anthropogenic in origin, as there are no known naturally occurring trifluoromethylated compounds. A second long-lived fluorinated acetic acid, chlorodifluoroacetic acid (CDFA) has recently been identified in rain water². Direct evidence of a source responsible for its production has yet to be established, although CFC-113 (CF₂ClCFCl₂) may contribute.

Agrochemicals, anaesthetics and fluoropolymers are the three main, distinct, categories into which environmental emissions of fluorine-containing anthropogenic organic compounds can be subdivided. Both agrochemicals¹⁴ and anaesthetics⁸ have been shown to produce TFA, but they can be eliminated as major contributors due to the quantities that are currently being emitted into the atmosphere⁸. Fluoropolymers, such as polytetrafluoroethylene (PTFE), are potential sources, due to their heavy usage in urban and industrial areas. In 1988 the average annual global consumption of fluorinated polymers was 40,000 tonnes (ref. 15) and by 1997 this figure had risen by almost 220% based upon sales¹⁶, with a projected annual increase of 7%.

In order to produce the CF₃– or CF₂Cl– unit required for the formation of haloacetic acid, the polymer would have to undergo decomposition and subsequent rearrangement. Fluoropolymers are being designed and modified specifically for use in areas of high thermal stress, such as ovens, cookware, industrial and car engines, heat exchangers, high-temperature circuits and a wide variety of other thermal applications. In 1997, 1% of all polymers were used in areas of elevated temperature, a grouping composed primarily of fluoropolymers¹⁷. The subcategory of plastics known as engineering plastics, in particular fluoropolymers, operate at the extreme of polymers' temperature performance¹⁷. PTFE, for instance, will endure 260 °C for ~2.3 years (ref. 18) until failure due to degradation¹⁹. Fluoropolymers are, largely, either recycled or degrade *in situ*¹⁹, resulting in research on the toxicity of the decomposition products²⁰. The onset of thermal degradation of fluoropolymers is known to initiate cleavage of the backbone and

Table 1 Positively identified species produced in the thermal decomposition of fluoropolymers

Polymer	Thermal product identified	Per cent produced
PTFE	TFE*†	–
	HFP*†	10.8§
	TFA*†	7.8‡§
	c-OFB†	–
	CF ₃ (CF ₂) _n COOH†	>0.01‡
	CF ₃ O(CF ₂) _m COOH†	–
	DFA†	>0.01‡
	MFA†	>0.01‡
OPTFE	CTFE*†	–
	CPFP*†	13.1§
	CDFA*†	9.5‡§
	TFA*†	>0.1‡§
	DCHB*†	–
	DCHFGB*	–
	TCTFE*	–
	1,3-DCTFP*†	–
	1,1,3-TCTFP*†	–
	CCl _x F _y (CCl _{x-1} F _{y-1}) _n COOH†	–
ECTFE	TFA*†	6.3‡§
	CDFA*†	7.2‡§
	HFP*	–
	CPFP*	–
PFEPE	TFA*†	2.5*†
	HFP*	–

A dash indicated that the analyte was positively identified but not quantified. Acronyms used which are not in the body of the text: MFA, monofluoroacetic acid; DFA, difluoroacetic acid; FDCA, fluorodichloroacetic acid; DCFP, dichloroperfluoropentanoic acid; DCFB, dichloroperfluorobutanoic acid; DCHFGB, 1,2-dichlorohexafluorocyclobutane; TCTFE, 1,1,2-trichloro-1,2,2-trifluoroethane; 1,3-DCTFP, 1,3-dichlorotetrafluoropropene; 1,1,3-TCTFP, 1,1,3-trichlorotrifluoropropene; ECTFE, ethylene-chlorotrifluoroethylene; PFEPE, polytetrafluoroethylene-co-tetrafluoroethylene perfluoropropylether. For the long-chain acids, $n = 1-12$, $m = 1-7$, $x = 0-2$, $y = 1-3$, $z = 1-2$, $p = 9-13$.

* Identified by NMR.

† Identified by MS.

‡ Quantified by NMR.

§ Quantified by MS.

subsequent rearrangement to produce significant amounts of trifluoromethylated species²⁰. The incineration, or combustion, of fluoropolymers has previously been proposed as a source of TFA⁸ and CDFA². These processes differ from thermolysis in that a source of fuel is used in order to purposefully evoke complete decomposition. Furthermore, it is unlikely to yield environmentally significant levels of TFA, or TFA precursors, due to the high temperatures and oxidizing conditions used, which would result in the cleavage of most carbon–fluorine bonds. On the other hand, low-temperature burning of domestic waste, which is an important source of polychlorinated dioxins and furans to the atmosphere²¹, may analogously be an important source of fluoroacids.

We aimed to show the possibility of direct and indirect production of haloacetic acids through the thermolysis of fluorinated polymers, and to investigate the production of other perhalogenated compounds that may be of environmental significance. We performed thermolysis of the fluoropolymer by placing the pure polymer in a quartz tube that was connected to a continuous stream of air. The tube was subsequently heated to a temperature of 500 °C. The experiment was also conducted isothermally at 360 °C for PTFE, a temperature that is used routinely in product sintering. The gases were collected as they left the reaction vessel, and immediately analysed. The temperature and conditions are not dissimilar to those that might be found in the burning of domestic wastes²¹.

Tetrafluoroethene (TFE), hexafluoropropene (HFP) and *cyclo*-octafluorobutane (*c*-OFB) were the main gases produced upon thermolysis of the pure fluorinated polymers and of the commercially available products tested (Table 1); CPTFE (chloropolytrifluoroethylene) analogously yielded chlorotrifluoroethene (CTFE), chloropentafluoropropene (CPFP) and 1,2-dichlorohexafluorocyclobutane (DCHB). HFP has the potential to react with OH radicals in the troposphere to produce TFA (100% conversion)¹¹. The reaction kinetics of CPFP with OH radicals are expected to be similar to that of HFP, based upon its reaction

with other radicals²² and the behaviour of similar molecules¹¹ producing CDFA in the troposphere. As well as PTFE and CPTFE, fluoropolymers ECTFE and PFEPE were also tested; see Table 1.

A large, previously unidentified, class of thermolysis compounds, perhalogenated acids, is also shown in Table 1. TFA and CDFA were the main acids to be observed in the thermolysis of the fluoro- and chlorofluoropolymers, while other longer-chain perhalogenated acids were also identified.

We propose a mechanism for the thermal degradation of PTFE, primarily involving the reaction of a CF₂ carbene unit, for the production of the main compounds observed (Fig. 1). This mechanism is supported by key products observed by several other workers (see Fig. 1 legend) and the additional products observed in this investigation. CPTFE and other polyfluorinated polymers have not been studied in nearly the same depth. We suggest that the mechanism for their degradation is similar, based upon the distribution of related products observed in this investigation.

Fluoropolymers are also incorporated into commercial products that may be exposed to high temperatures (such as the teflonized engine additive Slick 50, teflonized frying pans and surgical syringes). By heating these commercial products, we found that the incorporated fluoropolymers decomposed in such a way as to produce the same products as the pure polymers, that is, TFA and/or CDFA along with numerous other perhalogenated acids. Although the significance of any single source may be small, the incorporation of fluorinated polymers into commercial materials continues to increase, and could result in the accumulation of these compounds in the environment.

These results may explain the high TFA concentrations currently observed in precipitation (especially in urban areas), and the presence of other chlorofluoroacids, such as CDFA. They may also help to account for the seemingly contradictory reports that, in some instances, urban areas have higher concentrations of TFA in comparison with remote locations, while others have reported that the

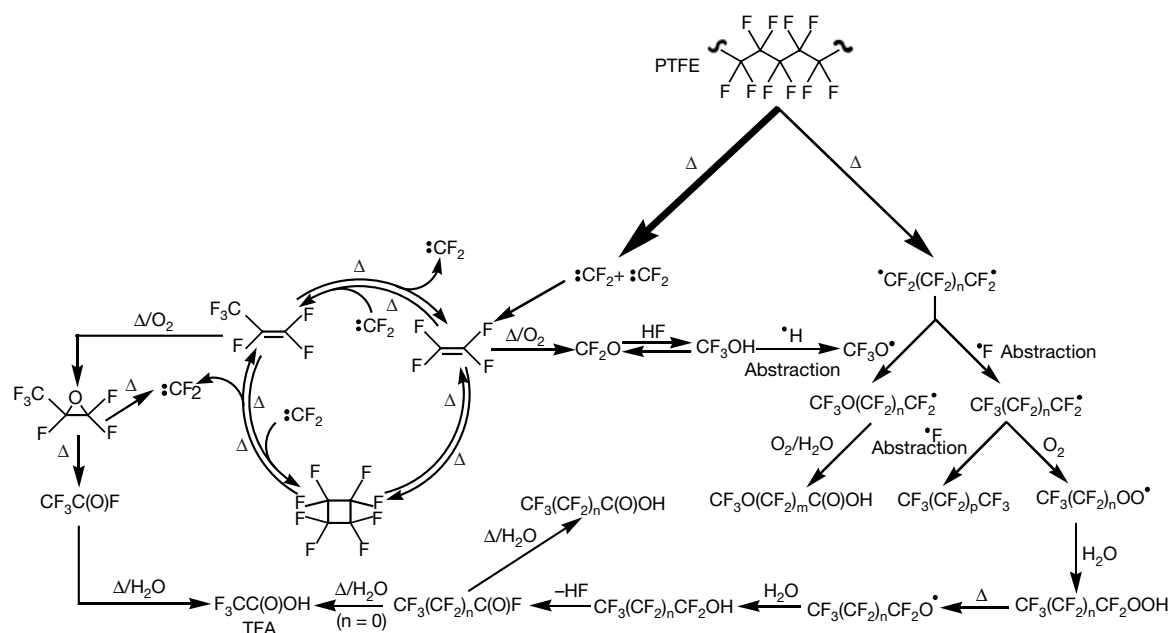


Figure 1 Proposed significant pathways in the thermal decomposition of fluoropolymers. The explicit major pathways for the production of TFA (trifluoroacetate) are shown. Evidence for the mechanism was obtained from this work and from previous observations^{19,20,25,28,29}. As indicated by the bold arrow, the most significant step in the thermal decomposition is the formation of carbene radicals. Longer-chain diradicals are also formed, which can undergo fluorine abstraction or reaction with carbonyl fluoride to

produce radical fluorinated alkanes ($p = 0-4$) or ethers. These radicals then react with constituents present in the air, oxygen and trace amounts of water, to form perfluorinated acids ($n = 0-12$, $m = 1-7$), the yield being inversely proportional to the number of carbon atoms in the chain. The distribution of product yield depends on temperature and the composition of the atmosphere. PTFE, polytetrafluoroethylene; Δ , heat.

concentrations are similar in both regions. These observations can be attributed to haloacetic acids being produced in two ways: directly through thermolysis of the polymer, and then undergoing short-range transport; and also indirectly, from the reaction of OH radicals with propenes, which may allow for longer-range transport. (We note that the tropospheric life-time of HFP is 9 days, based on its reaction kinetics with OH radicals; ref. 11). It has also been suggested that if HFP is produced from fluoropolymers, even to a small degree, that this would be a significant environmental burden of TFA¹¹.

Our hypothesis is that thermal degradation of fluoropolymers can lead to the formation of TFA. We have used environmental modelling to predict the concentration of TFA that is expected in Toronto rain. (Details of the model and the calculation are given in Supplementary Information.) A two-box environmental model was used; the smaller box (representing Toronto and the surrounding atmosphere) was contained within the larger box, representing North America. We first considered the flux of fluoropolymer-derived TFA to the atmosphere; this process was assumed to occur in Toronto and in North America as a whole. We then considered the flux of TFA into and out of the Toronto 'box'. After consideration of all other major fluxes, the average TFA concentrations in Toronto rainfall was estimated. The main uncertainty in these calculations is the percentage of fluoropolymers that ultimately undergo thermolytic degradation. We have used a value of 0.1% per year of the lowest estimated amount of fluoropolymer present in North America. This value is conservative due to the prevalent use of sintering—the pre-application heating of PTFE—involving tem-

peratures (360–382 °C) that our work has shown causes slow thermal breakdown. Further, the reported application of these polymers in areas of high thermal stress exceeds this value¹⁷. Finally, known indirect production of TFA through atmospheric oxidation of a significant thermal product, HFP, was not included within the model¹¹.

The model shows that there is the potential for a TFA concentration of $\sim 21 \text{ ng l}^{-1}$ in Toronto rainwater as a direct result of fluoropolymer thermolysis. These results are in accordance with those which have been experimentally observed²; furthermore, they are close to the level of urban enrichment recently observed in a study in California²³.

From an environmental-fate perspective, thermal degradation of the polymers produces monomeric units that will degrade in the troposphere to CO_2 (ref. 24), or halogenated propenes^{25,26} (Fig. 2) that degrade to produce long-lived haloacetic acids¹¹. Haloacetic acids and perhalogenated acids produced directly from the polymer are expected to be terrestrially deposited through wet and dry deposition⁴. Several perhalogenated acids, such as perfluorooctanoic acid and perfluorodecanoic acid, have been shown to act as peroxisome proliferators and as inhibitors of gap junctional intercellular communication⁶. (We note that perfluorooctanoic acid uses are being phased out by the manufacturer due to environmental persistence and pervasiveness⁷.) CFCs were produced through the decomposition of chlorofluoropolymers and may migrate to the stratosphere causing an effect on the ozone layer. The short-chain perfluorinated alkanes and cycloalkanes produced have an estimated tropospheric half-life of greater than 2,000 years,

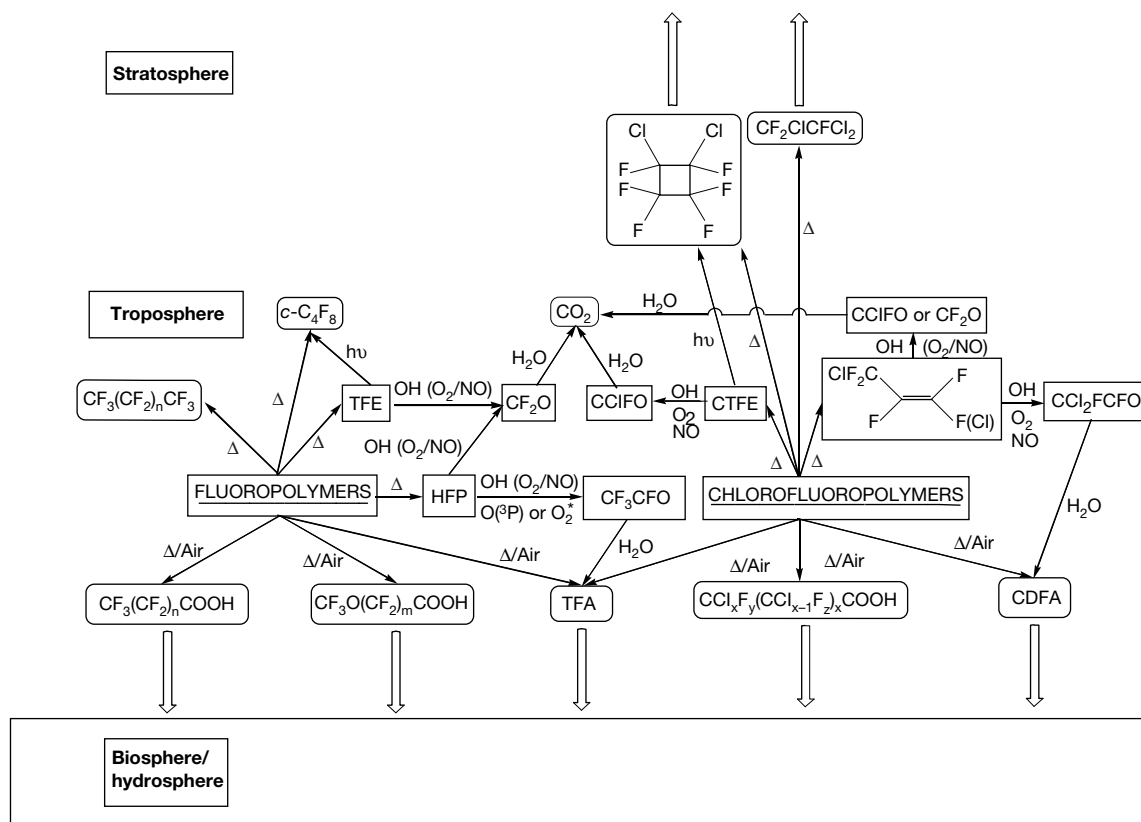


Figure 2 Proposed environmental reaction pathways for the thermal degradation of fluoro- and chlorofluoro- polymers. Rectangular boxes represent environmentally transient species ($t_{1/2} < 10$ years) and important environmental impacts. Reaction processes are either proposed on the basis of this study, or are based on refs 11, 20, 27, 30, 31. A significant concentration of perhaloacids ($x = 0-2$, $y = 1-3$, $z = 1-2$, $n = 1-12$, $m = 1-7$) are produced, primarily TFA (trifluoroacetate) and CDFA (chlorodifluoro-

acetic acid) (1–10 wt%). Tropospheric oxidation of the perhalopropenes also predominantly leads to, or is expected to lead to, the formation of TFA and CDFA. The thermal decomposition of the chlorofluoropolymers leads to the production of saturated chlorofluorocarbons (CFCs). TFE, tetrafluoroethylene; HFP, hexafluoropropene; CTFE, chlorotrifluoroethylene.

and have been proposed to effect global climate by acting as greenhouse gases²⁶.

We have shown that the thermolysis of fluorinated polymers results in the production of a range of compounds, most of which are extremely persistent, and whose environmental impact has yet to be fully assessed. Further studies are needed to quantify the actual emissions of these compounds from various sources. □

Methods

General thermolysis procedure

The polymer of interest (2 g) was placed in a quartz boat that was then inserted into a quartz tube (60 cm), and placed in an oven. Air was continually passed over the sample (~21 min⁻¹). The oven temperature was then increased to 500 °C at a rate of 11 °C min⁻¹. The evolved thermolysis materials were collected as described below. Product distribution was also confirmed at the lower isothermal temperature of 360 °C.

Collection procedures

For ¹⁹F NMR analysis of the volatile organics, the products were collected in a round-bottomed flask at -78 °C. The flask was then cooled to -196 °C, transferred to a vacuum line (~16 torr), and evacuated. An NMR tube containing 700 µl of CD₃CN cooled to -196 °C was connected in series with the flask. The flask was then allowed to warm to room temperature and the products transferred to the NMR tube. After one hour, the NMR tube was sealed, allowed to warm to room temperature, and an NMR spectrum collected immediately. Positive identification of products was done by comparing the spectra with those of authentic samples.

Analyses

For the analysis of perhalo-acids, the polymer decomposition off-gases were bubbled through an aqueous Na₂CO₃ solution (pH 10.2). For determination of quantities by ¹⁹F NMR, 500 µl of this solution was added to an NMR tube containing 500 µl of CD₃CN spiked with a known concentration of the internal standard, trifluoromethoxyacetaldehyde. Quantities were determined by comparison to an external calibration using a method similar to that of ref. 13. For identification and quantification using gas chromatography-mass spectrometry (GC-MS), aqueous samples were added to ethyl acetate containing 2,4-difluoroaniline and dicyclohexylcarbodiimide to produce the acid anilide in a method similar to that of ref. 2.

To measure the quantities of volatile gases produced in the polymer decomposition, the off-gases were collected in a glass vessel of known volume at -78 °C. After decomposition was complete, the vessel was sealed and allowed to warm to room temperature. Aliquots were removed using a gas-tight syringe and directly analysed using GC-MS. Quantification was performed by comparison to an external calibration.

Blanks consisting of all reagents were run to ensure that products observed were produced as a direct result of polymer decomposition.

Received 15 November 2000; accepted 22 May 2001.

- Franklin, J. The atmospheric degradation and impact of 1,1,1,2-tetrafluoroethane (hydrofluorocarbon 134a). *Chemosphere* **27**, 1565–1601 (1993).
- Martin, J. W. *et al.* Detection of chlorodifluoroacetic acid in North American precipitation and surface waters; a possible product of fluorocarbon degradation. *Environ. Sci. Technol.* **34**, 274–281 (2000).
- Boutonnet, J. C. *et al.* Environmental risk assessment of trifluoroacetic acid. *Hum. Ecol. Risk Assess.* **5**, 59–124 (1999).
- Cahill, T. M., Thomas, C. M., Shwarzbach, S. E. & Seiber, J. N. Accumulation of trifluoroacetate in seasonal wetlands in California. *Environ. Sci. Technol.* **35**, 820–825 (2001).
- Berends, A. G., Boutonnet, J. C., De Rooij, C. G. & Thompson, R. S. Toxicity of trifluoroacetate to aquatic organisms. *Environ. Toxicol. Chem.* **18**, 1053–1059 (1999).
- Upham, B. L., Deocampo, N. D., Wurl, B. & Trosko, J. E. Inhibition of gap junctional intercellular communication by perfluorinated fatty acids is dependant on the chain length of the fluorinated tail. *Int. J. Cancer* **78**, 491–495 (1998).
- Brown, D. & Mayer, C. E. 3M to pare Scotchguard products. *Wash. Post* **A01** (17 May 2000).
- Jordan, A. & Frank, H. Trifluoroacetate in the environment. Evidence for sources other than HFC/HCFCs. *Environ. Sci. Technol.* **33**, 522–527 (1999).
- Berg, M. *et al.* Concentrations and mass fluxes of chloroacetic acids and trifluoroacetic acid in rain and natural waters in Switzerland. *Environ. Sci. Technol.* **34**, 2675–2683 (2000).
- Wujcik, C. E., Cahill, T. M. & Seiber, J. N. Determination of trifluoroacetic acid in 1996–1997 precipitation and surface waters in California and Nevada. *Environ. Sci. Technol.* **33**, 1747–1751 (1999).
- Mashino, M. *et al.* Atmospheric chemistry of CF₃CF = CF₂: kinetics and mechanism of its reaction with OH radicals, chlorine atoms, and ozone. *J. Phys. Chem. A* **104**, 7255–7260 (2000).
- Von Sydow, L. M. *et al.* Natural background levels of trifluoroacetate in rain and snow. *Environ. Sci. Technol.* **34**, 3115–3118 (2000).
- Scott, B. F. *et al.* Haloacetic acids in Canadian lake waters and precipitation. *Environ. Sci. Technol.* **34**, 4266–4272 (2000).
- Ellis, D. A. & Mabury, S. A. The aqueous photolysis of TFM and related trifluoromethylphenols. An alternate source of trifluoroacetic acid in the environment. *Environ. Sci. Technol.* **34**, 632–637 (2000).
- Feiring, A. E. in *Organofluorine Chemistry: Principles and Applications* (eds Banks, R. E., Smart, B. E. & Tatlow, J.) 348 (Plenum, New York, 1994).

- Holloway, J. H. Fluorine, high-tech element for the next century. *J. Fluor. Chem.* **104**, 3–4 (2000).
- Johns, K. & Stead, G. Fluoroproducts—the extremophiles. *J. Fluor. Chem.* **104**, 5–18 (2000).
- Imbalzano, J. F. Combat corrosion with fluoroplastics and fluoroelastomers. *Chem. Eng. Prog.* **30**, 69–73 (1991).
- Simon, C. M. & Kaminsky, W. Chemical recycling of polytetrafluoroethylene by pyrolysis. *Polym. Degrad. Stab.* **62**, 1–7 (1998).
- Arito, H. & Soda, R. Pyrolysis products of polytetrafluoroethylene and polyethylenepropylene with reference to inhalation toxicity. *Ann. Occup. Hyg.* **20**, 247–255 (1977).
- Lemieux, P. M., Lutes, C. C., Abbott, J. A. & Aldous, K. M. Emissions of polychlorinated dibenzo-*p*-dioxins and polychlorinated dibenzofurans from the open burning of household waste in barrels. *Environ. Sci. Technol.* **34**, 377–384 (2000).
- Paleta, O., Cirkva, V., Budkova, Z. & Bohm, S. Radical addition reactions of fluorinated species part 6. Regioselectivity of the addition of nucleophilic radicals to halogenopropenes and evidence for a steric effect of the chlorine substituent. *J. Fluor. Chem.* **86**, 155–171 (1997).
- Cahill, T. M. & Seiber, J. N. Regional distribution of trifluoroacetate in surface waters downwind of urban areas in Northern California, U. S. A. *Environ. Sci. Technol.* **34**, 2909–2912, (2000).
- Acerboni, G., Jensen, N. R., Rindone, B. & Hjorth, J. Kinetics and products formation of the gas-phase reactions of tetrafluoroethylene with OH and NO₃ radicals and ozone. *Chem. Phys. Lett.* **309**, 364–368 (1999).
- Atkinson, B. & Atkinson, V. A. The thermal decomposition of tetrafluoroethylene. *J. Chem. Soc.* **32**, 2086–2094 (1957).
- Rizvi, M., Munir, A., Zulfiqar, S. & Zulfiqar, M. Pyrolysis gas-chromatographic studies of homopolymers and copolymers of chlorotrifluoroethylene with vinylidene fluoride and methyl methacrylate. *J. Therm. Anal.* **45**, 1597–1604 (1995).
- Ravishankara, A. R., Solomon, S., Turnipseed, A. A. & Warren, R. F. Atmospheric lifetimes of long-lived halogenated species. *Science* **259**, 194–199 (1993).
- Buravtsev, N. N. & Kolbanovsky, Y. A. Intermediates of thermal transformations of perfluoro-organic compounds. New spectral data and reactions. *J. Fluor. Chem.* **96**, 35–42 (1999).
- Schneider, W. F., Wallington, T. J. & Huie, R. E. Energetics and mechanism of decomposition of CF₃OH. *J. Phys. Chem.* **100**, 6097–6103 (1996).
- Hynes, R. G., Mackie, J. C. & Masri, A. R. Shock tube study of the oxidation of C₃F₆ by N₂O. *J. Phys. Chem. A* **103**, 5967–5977 (1999).
- Teruel, M. A., Taccone, R. A. & Lane, S. I. Discharge flow-chemiluminescence study of the rate coefficients for O(³P) + CF₂ = CCl₂ and O(³P) + CF₃CF = CF₂ reactions at 298 K. *Int. J. Chem. Kinet.* **31**, 867–872 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank H. Sun, I. Mannes, M. Ginzburg, D. McIntosh, D. Mathers, W. Yoo and E. Novata for assistance with experimental procedures and the use of equipment, and F. Wania for assistance in environmental modelling. This work was supported by a Strategic Project Grant from the Natural Science and Engineering Research Council of Canada (NSERC); we thank Perkin Elmer for their equipment donation to the University of Toronto ANALEST facility.

Correspondence and requests for materials should be addressed to S.A.M. (e-mail: smabury@chem.utoronto.ca).

The role of microbial mats in the production of reduced gases on the early Earth

Tori M. Hoehler, Brad M. Bebout & David J. Des Marais

Exobiology Branch, NASA Ames Research Center, MS 239-4, Moffett Field, California 94035-1000, USA

The advent of oxygenic photosynthesis on Earth may have increased global biological productivity by a factor of 100–1,000 (ref. 1), profoundly affecting both geochemical and biological evolution. Much of this new productivity probably occurred in microbial mats, which incorporate a range of photosynthetic and anaerobic microorganisms in extremely close physical proximity^{2,3}. The potential contribution of these systems to global biogeochemical change would have depended on the nature of the interactions among these mat microorganisms.

Here we report that in modern, cyanobacteria-dominated mats from hypersaline environments in Guerrero Negro, Mexico, photosynthetic microorganisms generate H_2 and CO —gases that provide a basis for direct chemical interactions with neighbouring chemotrophic and heterotrophic microbes⁴. We also observe an unexpected flux of CH_4 , which is probably related to H_2 -based alteration of the redox potential within the mats. These fluxes would have been most important during the nearly 2-billion-year period during which photosynthetic mats contributed substantially to biological productivity⁵—and hence, to biogeochemistry—on Earth. In particular, the large fluxes of H_2 that we observe could, with subsequent escape to space, represent a potentially important mechanism for oxidation of the primitive oceans and atmosphere.

We measured fluxes and *in situ* concentrations of H_2 and CO in two microbial mat communities representing different modes of growth. In a hypersaline (salinity, 95‰) pond, the cyanobacterium *Microcoleus chthonoplastes* constructs cohesive mats that are 4–5 cm thick. These mats are permanently submerged, rarely experience physical disruption, and are approximately steady-state with respect to growth/decomposition⁶. These conditions seem to foster the growth of anaerobic microbial populations, which contribute prominently to the overall biogeochemistry of these subtidal communities⁷. A second group of mats, located on an intertidal sand flat, is constructed predominantly by *Lyngbya* spp. cyano-

bacteria. Frequent physical disruption, combined with alternating periods of desiccation/aeration and tidal flooding with lagoon water, forces these *Lyngbya* communities to exist in a perpetually pioneering mode of more rapid growth, in which the requirement for nitrogen fixation is enhanced⁸. Anaerobic processes make a relatively small contribution to the overall biogeochemistry of these mats, suggesting that the fluctuating environmental conditions may inhibit the development of substantial anaerobic microbial populations. A fibrous surface texture efficiently traps and retains photosynthetically generated gas bubbles, so gas production in these intertidal mats has a more lasting impact on surface communities than in subtidal (*M. chthonoplastes*) mats, where the smooth surface does not retain bubbles.

Both mat communities exhibit production of carbon monoxide that is clearly tied to light intensity (Fig. 1a). The partial pressure of CO (p_{CO}) in entrapped gas bubbles on the surface of intertidal mats is up to 4.5 Pa during daylight hours, with levels dropping by a factor of 10–20 at night (Fig. 1b). In subtidal mats (the thickness and structure of which allow for better resolution in vertical profiles), this CO production is clearly associated with the photic surface layer (Fig. 2a) and does not occur when samples are incubated in the dark. CO production rates are 75 times higher in untreated mats than in those amended with 3-(3,4-dichlorophenyl)-1;1-dimethylurea (DCMU) (a specific inhibitor of photosystem II), indicating a link to oxygenic photosynthesis. Collectively, these factors suggest that CO production is closely connected to cyanobacterial photosynthesis. To our knowledge, production of CO by cyanobacteria has not previously been reported in culture or in the environment.

Mats also generate large quantities of H_2 within the surface layer, in a daily pattern that is opposite that of CO production. In bubbles on the surface of *Lyngbya* mats, the H_2 partial pressure (p_{H_2}) varies by four orders of magnitude during the course of a 24-h diel cycle—daytime levels are held below 1 Pa, followed by a steady rise to 10 kPa ($\sim 10\%$ H_2) at night (Fig. 1c). Values of p_{H_2} are highest in the upper 2 mm of the mat, where cyanobacteria are active (Fig. 2b), and mats deprived of sunlight during the day exhibit a 4 times smaller flux of H_2 at night—indicating that H_2 production in these mats is closely dependent on cyanobacterial photosynthesis. However, the highest concentrations of H_2 develop in the dark, and H_2 fluxes are stimulated by a factor of up to 150 by anoxic conditions (Fig. 1c; Table 1); cyanobacterial photosynthesis can be only indirectly responsible for this activity. We suggest that carbon and reducing power fixed by cyanobacteria during the day fuel night-time production of H_2 by the O_2 -sensitive processes of nitrogen fixation and fermentation^{9–12}. Aerobic or phototrophic oxidation of H_2 may

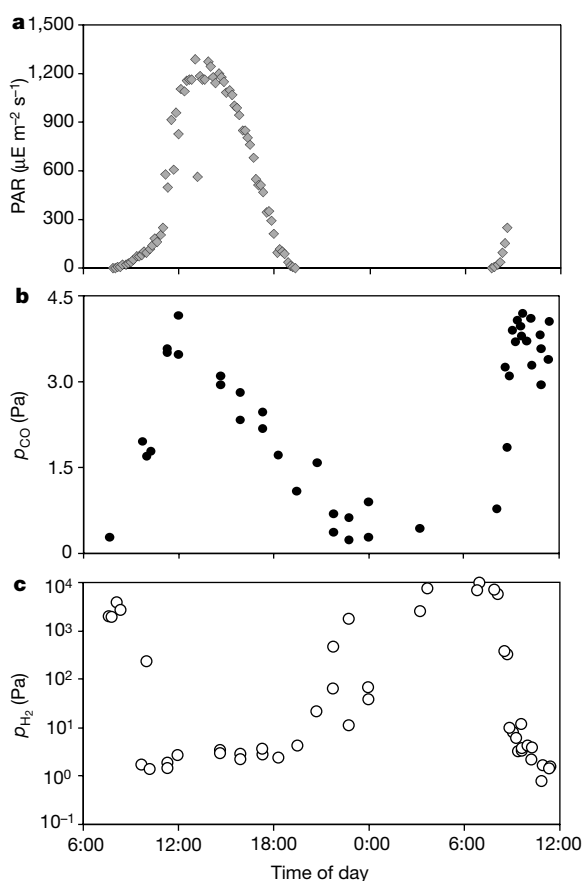


Figure 1 Light-driven gas chemistry in the surface layer of intertidal mats dominated by *Lyngbya* spp. cyanobacteria. **a**, Photosynthetically available radiation (PAR) versus time of day at the surface of the overlying water. **b**, Partial pressure of CO (p_{CO}) in small ($\sim 1\text{-}\mu l$) gas bubbles formed at the surface of the mat via photosynthesis. **c**, Partial pressure of H_2 (p_{H_2}) in the same bubble samples. We note the logarithmic scale on the partial-pressure axis, and that H_2 varies by four orders of magnitude during the diel cycle. See Methods for details of gas bubble analysis.

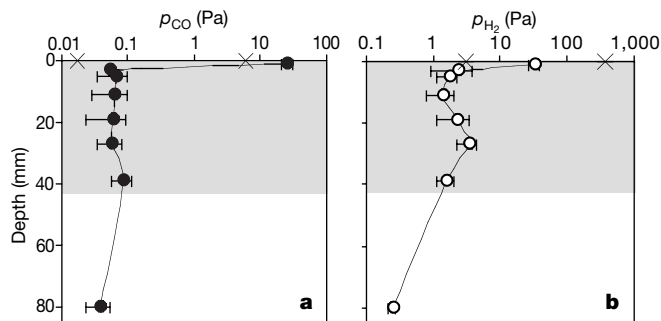


Figure 2 Depth distribution of partial pressures of CO (**a**) and H_2 (**b**) in *M. chthonoplastes* mats. The cohesive mat layer is indicated by grey shading. Crosses on the partial-pressure axis reflect the daily range in bubbles captured from the mat surface (see Fig. 1). Error bars represent ± 1 s.d. about the mean of triplicate samples. See Methods for details of depth profile determinations.

Table 1 Measured gas fluxes from microbial mats

Source	Gas		
	Carbon monoxide ($\mu\text{mol m}^{-2} \text{d}^{-1}$)	Hydrogen ($\mu\text{mol m}^{-2} \text{d}^{-1}$)	Methane ($\mu\text{mol m}^{-2} \text{d}^{-1}$)
Intertidal, anoxic	3.1 ± 0.1	$15,100 \pm 4,700$	0.6 ± 0.003
Intertidal, oxic	3.0 ± 0.4	96 ± 90	0.3 ± 0.07
Subtidal, anoxic	5.4 ± 0.7	83 ± 17	17.3 ± 0.5
Subtidal, oxic	5.5 ± 0.8	58 ± 9	19.5 ± 1.1

Comparison of gas fluxes from intertidal (*Lyngbya*) and subtidal (*M. chthonoplastes*) microbial mat communities. Fluxes of CO and H₂ are not constant over the diel cycle, but instead follow a pattern corresponding to the daily rise and fall of corresponding partial pressures in the surface layer: CO flux occurs exclusively during daylight hours, while H₂ fluxes occur in the dark. The values above reflect the total fluxes measured over 24 h. The average of two treatments (± 1 s.d.) is given.

also be important in regulating the daily pattern of concentrations and fluxes.

Both mats (but particularly the subtidal ones) exhibit an unexpected production of CH₄, which is probably related to cyanobacterial production of H₂. Typically, high sulphate concentrations (50–80 mM in the mats described here) allow sulphate-reducing bacteria to efficiently out-compete CH₄-producing Archaea for the common substrate H₂, rendering methanogenesis thermodynamically unfavourable¹³. Within the cohesive mat layer, however, H₂ concentrations are at least an order of magnitude higher than in the unconsolidated sediments beneath (Figs 2b, 3), effectively relieving any thermodynamic inhibition of methanogenesis. CH₄ production is stimulated in this zone of increased H₂ concentrations, but does not occur in the low-H₂, unconsolidated sediments beneath (Fig. 3). This is one of potentially many instances in which the production of H₂ may influence individual metabolism, microbial population structure, and system-level chemistry. A broad spectrum of anaerobic microbial processes utilize H₂ and CO (ref. 4), and are subject to thermodynamic control in direct analogy to methanogenesis^{14–17}. By virtue of this control, cyanobacterial production of H₂ and CO may have a broad-based effect on the microbial ecology and resultant system-level biogeochemistry of microbial mat communities.

Fluxes of H₂ and CH₄ to the environment differ substantially between the two mats examined, in a pattern that reflects the relative importance of anaerobic microbial communities within the mat. In intertidal *Lyngbya* mats, where frequent aeration and desiccation appears to hinder the development of a substantial anaerobe population, CH₄ fluxes are 30–40 times lower than in subtidal *M. chthonoplastes* mats, which harbour a well-developed population of anaerobes. These fluxes are little affected by oxic versus anoxic incubation conditions (Table 1), suggesting that the large difference between mats stems not from methane oxidation, but from differences in the size or activity level of methanogenic populations. An opposite effect is observed for H₂ flux. In intertidal mats,

Table 2 Possible gas fluxes on the early Earth

Gas	Source		
	Ancient geothermal	Intertidal mats	Subtidal mats
Carbon monoxide (Tmol yr ⁻¹)	0.03–0.36	0.05–0.5	0.014–0.14
Hydrogen (Tmol yr ⁻¹)	0.03–0.15	64–640	0.43–4.3
Methane (Tmol yr ⁻¹)	0.07–0.24	0.006–0.06	0.15–1.5
All reduced volatiles (Tmol yr ⁻¹)	2–20		

Comparison of possible gas fluxes from geological and biological sources on the early, pre-O₂ Earth. 'Ancient geothermal' fluxes are modern estimates, scaled up by 10³ to account for greater heat flux and crustal spreading rates on early Earth. All modern flux estimates except CO are from ref. 26. CO flux is estimated from the modern CO₂ flux²⁶ by assuming a CO:CO₂ ratio of 0.03 in modern volcanic emissions²¹. Measured fluxes from intertidal (*Lyngbya*) and subtidal (*M. chthonoplastes*) mats under anoxic conditions were expressed as a fraction of net primary productivity (NPP) in the mat sample, and then scaled up to an ancient global mat-based NPP equal to 10–100% of modern marine NPP (4,000 Tmol C yr⁻¹)²⁷. Mat-based fluxes on the early Earth would always reflect a combination of intertidal and subtidal inputs, and so would average the above estimates (for example, an equal mix of mat inputs would result in CO, H₂ and CH₄ fluxes of 0.032–0.32, 32.2–322 and 0.078–0.78 Tmol yr⁻¹, respectively).

cyanobacterial H₂ production is relatively unaffected by anaerobic processes, so the flux of H₂ is large: night-time release of H₂ represents about 16% of the reducing power fixed during the day (Table 1). In contrast, the well-developed anaerobe population in subtidal mats seems to substantially decrease H₂ fluxes in two ways. First, by decreasing the need for cyanobacterial nitrogen fixation and concomitant H₂ production through efficient remineralization of organic-matter nitrogen⁸; and second, by providing a large capacity for anaerobic consumption of H₂. Fluxes are much smaller than those from intertidal mats (Table 1), with only 0.2% of the reducing power fixed during daylight hours released as H₂ at night. This is in agreement with previous observations that mats with large attendant populations (as in *M. chthonoplastes* mats) could support large H₂ fluxes, but only when the activities of H₂-consuming sulphate-reducers were completely inhibited¹⁸. CO fluxes from the two communities are similar (Table 1), perhaps indicating that the photosynthetic source is little affected by the presence or absence of anaerobic populations in the mat.

The significance of CO, H₂ and CH₄ production by microbial mats is tied to the historical importance of cyanobacterial productivity on Earth. Before the advent of oxygenic photosynthesis, the productivity of the biosphere (and hence, the potential influence of biology on global geochemistry) must have been dependent entirely on the flux of reducing power from geological sources. The evolution of cyanobacteria, with their ability to use water as a source of reducing power, probably increased the productivity of Earth's biosphere by 2–3 orders of magnitude¹, with two key effects.

First, most of the reducing power available to non-photosynthetic microorganisms on Earth now came through physical association with cyanobacterial communities, making mats a major habit for microbial life on Earth for perhaps the next two billion years. One of the most important stages in the evolution of life on Earth—adaptation to environmental O₂—may thus have unfolded predominantly within the biogeochemical microenvironment of microbial mats^{19,20}. As we have seen in the case of methanogenesis, cyanobacterial production of CO and H₂ may strongly affect the mat microenvironment by altering the thermodynamics of microbial metabolism^{16,17}, the exchange of reducing power between functional groups^{14,15}, and, potentially, the population structure and spatial organization of the overall mat community. Collectively,

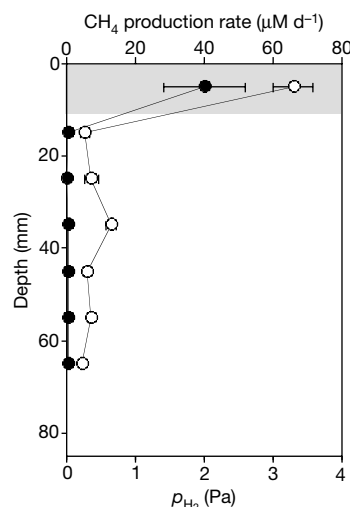


Figure 3 Depth distribution of H₂ partial pressure (open circles) and CH₄ production rate (filled circles). Mats were collected from the same subtidal pond as those in Fig. 2, but at a lower salinity and higher sedimentation rates, and were consequently thinner (1 cm versus 4–5 cm). The cohesive mat layer is indicated by grey shading. Error bars represent ± 1 s.d. about the mean of duplicate samples.

these factors establish the very matrix in which much of microbial evolution may have unfolded, so that cyanobacterial production of CO and H₂ should be viewed as a key influence on the evolutionary process during this critical period.

Second, the biosphere, for the first time, could compete substantially with geological sources as the major influence on the chemistry of the oceans and atmosphere. If the productivity attributable to microbial mats on early Earth was similar to that of modern marine ecosystems, even within an order of magnitude, virtually any aspect of their chemistry would have been significant on a global scale. The fraction of productivity that is transformed to CO and CH₄ in modern mats would, on the early Earth, scale to fluxes equal to, or significantly exceeding, those from geothermal sources (Table 2). The mat-driven flux of H₂ would have exceeded the geothermal flux by 2–4 orders of magnitude, and the flux of all reduced volatiles by an order of magnitude or more (Table 2)—a profound transmission of photosynthetic reducing power to the environment.

The very large flux of H₂ is interesting in a geochemical context, as a potential mechanism for oxidation of the Earth's surface. Net oxidation of the planet via oxygenic photosynthesis is only possible when the reducing power generated during the process is effectively removed from the system²¹. This removal is known to occur via burial of organic matter (reduced carbon) in deep reservoirs^{22,23}, but could also occur if mat-produced H₂ was lost to space²⁴—with a net contribution of one O₂ molecule for every two H₂ molecules lost. The carbon burial flux is estimated to contribute approximately 10 Tmol O₂ to the atmosphere on an annual basis²². By comparison, the mat-driven atmospheric flux of H₂, if completely lost to space, would correspond to a net O₂ flux of 0.2–320 Tmol (Table 2). Owing to the sensitivity of H₂ fluxes to O₂ tension (Table 1), this mechanism would have been most important during the period when atmospheric O₂ levels were low. Therefore, H₂ escape might have contributed substantially to the oxidation of Earth's large reservoirs of reduced iron and sulphur, which must have preceded actual atmospheric oxygenation.

Broadly speaking, the fluxes of CO, CH₄ and H₂ represent individual examples of what must be generally true: that even minor aspects of microbial mat chemistry, when amplified through the high productivity of oxygenic photosynthesis, must be considered as potentially profound influences on global biogeochemistry for more than half of Earth's history. □

Methods

Bubble analysis

Lyngbya mats were excised from a tidal flat in 1-cm-thick, 25×25 cm squares, placed in acrylic aquaria that were transparent to ultraviolet radiation, covered with 5 cm of water from the field site (salinity, 60‰), and incubated under natural solar irradiance and regulated temperature (20 °C, the approximate *in situ* average). Small (~1-μl) gas bubbles formed at the mat surface as a result of photosynthesis, and persisted through the dark period due to entrapment in a filamentous surface texture. These were collected by means of a 3-ml syringe with an inverted pipette tip lodged in the Luer fitting. When filled with water and submerged into the incubation tank, this device could be used to collect individual bubbles non-invasively, without atmospheric contamination, and pool them to a total volume of 5–25 μl. Gas collected in the plastic syringe was immediately subsampled with a volumetric gas-tight syringe for analysis by gas chromatography (GC).

Depth profiles

These were determined for two mats: thick (~5 cm) *M. chthonoplastes* mats taken from a low-sedimentation, permanently submerged, high-salinity (95‰) site (Fig. 2); and thinner (~1 cm) mats taken from the opposite end of the same pond under conditions of lower salinity (70‰) and higher sedimentation (Fig. 3). Previously described methods were utilized for determination of H₂ concentrations in sediments²⁵, with CO determined in the same analysis. Treatment of samples of *M. chthonoplastes* mats deviated from the published methodology in that: (1) incubations used 'disks' of mat 2 mm thick and 1.2 cm in diameter suspended in 4 ml of pond water; and (2) the uppermost depth intervals were incubated under natural solar irradiance and with an air headspace, while deeper intervals received an N₂ headspace and no light, each simulating *in situ* conditions. Calculations of CH₄ production rate were based on the increase in CH₄ concentrations over three successive samplings of the headspace gas in vials incubated for H₂ concentration determinations.

Flux experiments

Square sections of *Lyngbya* and *M. chthonoplastes* mats were excised and incubated as described above for 'bubble analysis', with the exception that the latter mats were incubated under pond water of 95‰ salinity and both mat types were covered with water to a depth of 10 cm. Stirred flux chambers (glass, 1.3 l internal volume) were emplaced and amended with a 120-ml gas headspace (pure N₂ for night-time 'anoxic' and all daytime treatments; 90 ml air + 30 ml O₂ for 'oxic' treatments). Incubations were for a minimum of 24 h, with the water and headspace in each chamber reset between the day and night periods. Gas in the chamber headspace was sampled directly through a septum using a volumetric gas-tight syringe.

Gas analyses

Determination of H₂ and CO concentrations in sampled gas was made in a single GC run with HgO detection (using a Trace Analytical RGA-3), always within 30 s of sampling. Determination of CH₄ concentrations, either immediately or on samples stored in glass vials for not more than 6 h, was by GC with Flame ionization detection.

Received 3 November 2000; accepted 17 May 2001.

- DesMarais, D. J. in *Geomicrobiology* (eds Banfield, J. & Nealson, K.) 429–445 (Mineralogical Society of America, Washington DC, 1997).
- Cohen, Y. et al. *Ancient Stromatolites and Microbial Mats* (Liss, New York, 1984).
- Stal, L. J. & Caumette, P. *Microbial Mats: Structure, Development and Environmental Significance* (Springer, Berlin, 1994).
- Conrad, R. Biogeochemistry and ecophysiology of atmospheric CO and H₂. *Adv. Microb. Ecol.* **10**, 231–283 (1988).
- Walter, M. R. (ed.) *Stromatolites* (Elsevier, Amsterdam, 1976).
- DesMarais, D. J. in *Advances in Microbial Ecology* (ed. Jones, J. G.) 251–274 (Plenum, New York, 1995).
- Canfield, D. E. & DesMarais, D. J. Biogeochemical cycles of carbon, sulfur, and free oxygen in a microbial mat. *Geochim. Cosmochim. Acta* **57**, 3971–3984 (1993).
- Bebout, B. M., Paerl, H. W., Bauer, J. M., Canfield, D. E. & DesMarais, D. J. in *Microbial Mats: Structure, Development, and Environmental Significance* (eds Stal, L. J. & Caumette, P.) 265–271 (Springer, Berlin, 1994).
- Newton, W., Postgate, J. R. & Rodriguez-Barrueco, C. *Recent Developments in Nitrogen Fixation* (Academic, London, 1977).
- Stal, L. J., Heyer, H., Bekker, S., Villbrandt, M. & Krumbein, W. E. in *Microbial Mats: Physiological Ecology of Benthic Microbial Communities* (eds Cohen, Y. & Rosenberg, E.) 255–276 (American Society for Microbiology, Washington DC, 1989).
- Fay, P. Oxygen relations of nitrogen fixation in cyanobacteria. *Microbiol. Rev.* **56**, 340–373 (1992).
- Stal, L. J. & Moezelaar, R. Fermentation in cyanobacteria. *FEMS Microbiol. Rev.* **21**, 179–211 (1997).
- Lovley, D. R. & Klug, M. J. Sulfate reducers can outcompete methanogens at freshwater sulfate concentrations. *Appl. Environ. Microbiol.* **45**, 187–192 (1983).
- Iannotti, E. L., Kafkewitz, D., Wolin, M. J. & Bryant, M. P. Glucose fermentation products of *Ruminococcus albus* grown in continuous culture with *Vibrio succinogenes*: Changes caused by interspecies transfer of H₂. *J. Bacteriol.* **114**, 1231–1240 (1973).
- Wolin, M. J. & Miller, T. L. Interspecies hydrogen transfer: 15 years later. *ASM News* **48**, 561–565 (1982).
- Dolfing, J. in *Biology of Anaerobic Microorganisms* (ed. Zehnder, A. J. B.) 417–468 (Wiley-Interscience, New York, 1988).
- Schink, B. in *Biology of Anaerobic Microorganisms* (ed. Zehnder, A. J. B.) 771–846 (Wiley-Interscience, New York, 1988).
- Skyring, G. W., Lynch, R. M. & Smith, G. D. in *Microbial Mats: Physiological Ecology of Benthic Microbial Communities* (eds Cohen, Y. & Rosenberg, E.) 170–179 (American Society for Microbiology, Washington DC, 1989).
- Fenchel, T. & Finlay, B. J. *Ecology and Evolution in Anoxic Worlds* (Oxford Univ. Press, Oxford, 1995).
- Nisbet, E. G. & Fowler, C. M. R. Archaean metabolic evolution of microbial mats. *Proc. R. Soc. Lond. B* **266**, 2375–2382 (1999).
- Holland, H. D. *Chemical Evolution of the Atmosphere and Oceans* (Princeton Univ. Press, Princeton, 1984).
- Berner, R. A. & Canfield, D. E. A new model for atmospheric oxygen over Phanerozoic time. *Am. J. Sci.* **289**, 333–361 (1989).
- DesMarais, D. J., Strauss, H., Summons, R. E. & Hayes, J. M. Carbon isotope evidence for the stepwise oxidation of the Proterozoic environment. *Nature* **359**, 605–609 (1992).
- Walker, J. C. G. *Evolution of the Atmosphere* (Macmillan, New York, 1977).
- Hoehler, T. M., Alperin, M. J., Albert, D. B. & Martens, C. S. Field and laboratory studies of methane oxidation in an anoxic marine sediment: Evidence for a methanogen-sulfate reducer consortium. *Glob. Biogeochem. Cycles* **8**, 451–463 (1994).
- Elderfield, H. & Schulz, A. Mid-ocean ridge hydrothermal fluxes and the chemical composition of the ocean. *Annu. Rev. Earth Planet. Sci.* **24**, 191–224 (1996).
- Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).

Acknowledgements

We thank Exportadora de Sal, S.A. de C.V. for access to their salt ponds and for logistical support, and S. Miller, P. Visscher and L. Jahnke for discussions. This work was supported by NASA's Astrobiology Institute and Exobiology Program, and by Ames Research Center Director's Discretionary Funds. T.M.H. was supported by a National Research Council fellowship.

Correspondence and requests for materials should be addressed to T.M.H. (e-mail: thoechler@mail.arc.nasa.gov).

Reproductive pair correlations and the clustering of organisms

W. R. Young^{*}, A. J. Roberts[†] & G. Stuhne^{*}

^{*} Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093-0230, USA

[†] Department of Maths & Computing, University of Southern Queensland, Toowoomba, Queensland 4352, Australia

Clustering of organisms can be a consequence of social behaviour, or of the response of individuals to chemical and physical cues¹. Environmental variability can also cause clustering: for example, marine turbulence transports plankton^{2–8} and produces chlorophyll concentration patterns in the upper ocean^{9–11}. Even in a homogeneous environment, nonlinear interactions between species^{12–14} can result in spontaneous pattern formation. Here we show that a population of independent, random-walking organisms ('brownian bugs'), reproducing by binary division and dying at constant rates, spontaneously aggregates. Using an individual-based model, we show that clusters form out of spatially homogeneous initial conditions without environmental variability, predator–prey interactions, kinesis or taxis. The clustering mechanism is reproductively driven—birth must always be adjacent to a living organism. This clustering can overwhelm diffusion and create non-poissonian correlations between pairs (parent and offspring) or organisms, leading to the emergence of patterns.

Figure 1 shows a simulation of the brownian bug model using an individual-based Monte Carlo approach (described below in Methods). Because the motion of a brownian bug is unaffected by a reproductive event, or by the proximity of other bugs, the spatial aspects of the model are only weakly coupled to the primitive biological ingredients of birth and death. Perhaps, then, it is surprising that continuum approximations do not represent the patches that form in the individual-based model shown in Fig. 1. The continuum approximation^{1,12} describes the evolution of large populations using advection–diffusion–reaction (ADR) equations for the concentration: $C(\mathbf{x}, t)dA$ is the expected number of organisms in the sample area dA surrounding the point $\mathbf{x}$ at time t . The ADR approximation of the brownian bug process is:

$$C_t = D\nabla^2 C + (\lambda - \mu)C \quad (1)$$

where λ is the birth rate, μ is the death rate and D is the diffusivity. However, in Fig. 1 death and birth are equally probable so $\lambda = \mu$, and equation (1) collapses to the diffusion equation. Then, as the initial concentration C_0 is uniform, the solution of equation (1) is $C(\mathbf{x}, t) = C_0$; this is not a good characterization of Fig. 1. This failure shows that equation (1), and other ADR approximations^{2–8,13–15}, do not capture the fluctuations exhibited by the individual-based model in Fig. 1.

In reality there is a fundamental difference between the locations of birth and death: deaths occur anywhere, but birth always occurs adjacent to a living organism. This asymmetry is not represented in the continuum approximation of equation (1), in which the birth rate and the death rate occur only in the combination $\lambda - \mu$. These reproductive pair correlations are a uniquely biological complication, with no analogue in the physical and chemical problems that served as the initial inspiration¹⁵ for biological ADR modelling. Careful derivations of the ADR approximation¹² make the assumption that on the length scale of the sample area dA individuals are distributed without correlations. This assumption enables us to use Poisson statistics to calculate averages. The brownian bug process realized in Fig. 1 shows that this Poisson hypothesis can fail in simple circumstances.

Of course, uniform concentration is the correct (but useless) answer if $C(\mathbf{x}, t)$ is defined by an ensemble average over many realizations: the concentration fluctuations in Fig. 1 disappear after an ensemble average because the brownian bug process is spatially homogeneous. However, as discussed above, the concentration is usually defined via a spatial average over dA . The optimal choice of sampling scale is important^{16–18} because in most circumstances only spatial averages are accessible to observation, whereas the ensemble average is most useful for theoretical purposes. If the spatial average of a single realization is approximated by the ensemble average of many realizations (for example, either molecularly or turbulently diffusing chemicals) we say that the system is self-averaging. The

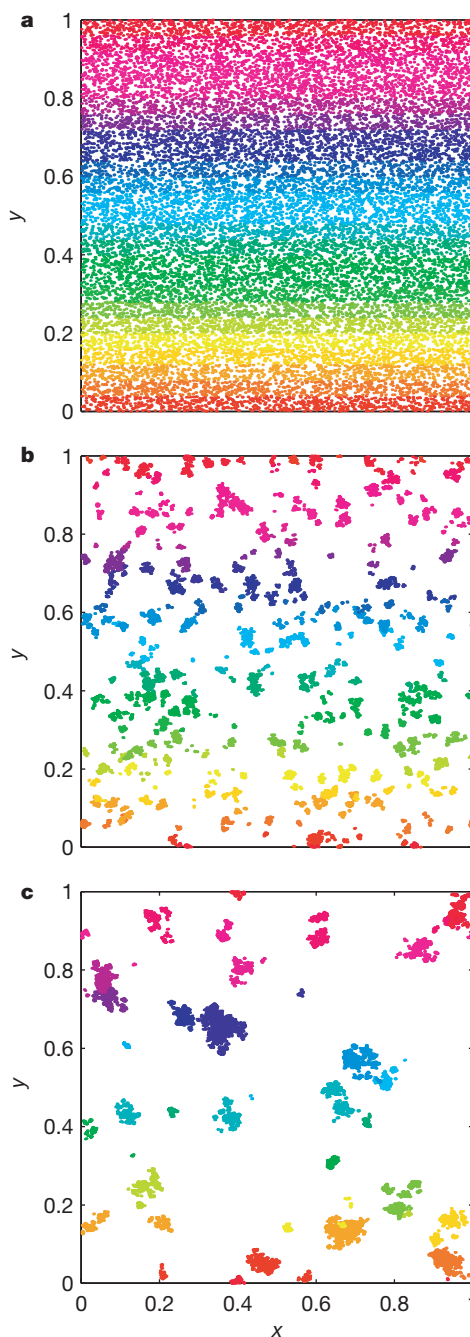


Figure 1 Distributions of brownian bugs at different times in a simulation with $\Delta = 10^{-3}$ and $N_0 = 20,000$. Each point is the position of a brownian bug in the (x, y) -plane. See Methods for the details of the simulation. **a**, Initial condition (a Poisson spatial distribution), with brownian bug colour coded by y -coordinate. **b**, Descendants of the original brownian bugs at $t = 100\tau$. **c**, As **b**, at $t = 1,000$.

disagreement between the solution $C(\mathbf{x}, t) = C_0$ of equation (1) and the realization in Fig. 1 indicates a failure self-averaging.

For oceanographic systems it is interesting to study the effects of large-scale random advection on the patches in Fig. 1. This is the simplest model of a planktonic species reproducing in a turbulent fluid. The assumption that the flow field is large scale means that the characteristic length of the velocity is much greater than the diameter of the emerging patches in Fig. 1b. In other words, we are in the 'Batchelor regime' in which the velocity field varies linearly with the separation $s(t)$ of a recently divided parent and offspring^{19–22}. In this case there is a timescale (related to the root mean square, r.m.s., strain rate of the velocity) but there is no length scale, other than $s(t)$. Consequently $s(t) \propto \exp(\gamma t)$, where γ^{-1} is the timescale mentioned above.

Because exponential pair separation is much faster than diffusive separation, we might expect that the patches formed by reproductive pair correlations would disappear. However, the simulation in Fig. 2, which uses a computationally tractable model of exponential pair separation²², shows that the patches persist as elongated filaments (Fig. 2b). This elongation is suggested by visualizations of nonreproducing contaminants in this same incompressible random velocity field²² (Fig. 2a). Without reproduction an initially uniform density of brownian bugs remains uniform (Fig. 2a): stirring alone does not create patches. But with reproduction, filamentary patches appear spontaneously (Fig. 2b). Thus, there is an essential difference between reproductive and nonreproductive contaminants.

To proceed with theory beyond equation (1) we notice that the constant ensemble-averaged concentration C is the first member of a

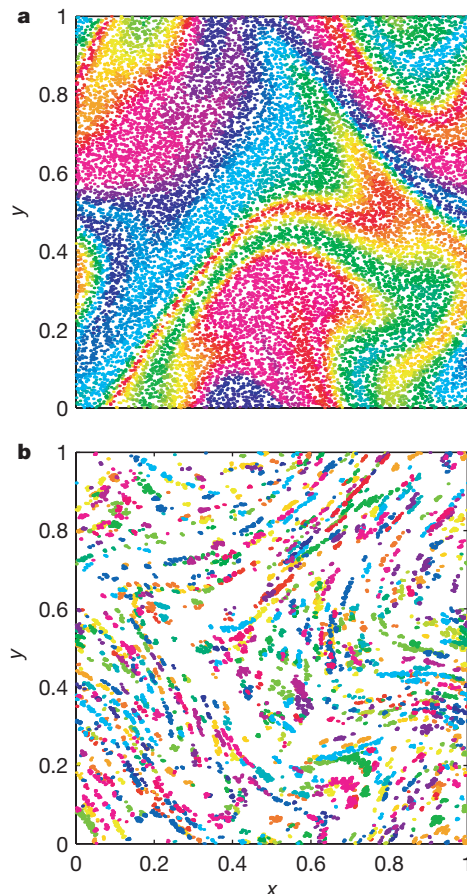


Figure 2 Effects of stirring on the brownian bug clustering ($\Delta = 0.001$, $N_0 = 20,000$, as in Fig. 1). Each point is the position of a brownian bug in the (x, y) -plane. **a**, Distribution at $t = 30\tau$ with $U/\tau/2 = 0.1$ and no birth or death. **b**, Distribution at $t = 1,000\tau$ and $U/\tau/2 = 0.1$.

hierarchy of distribution functions that describe the nonindependent probabilities of the positions of all brownian bugs²³. The second member of this hierarchy is the pair correlation function: $G(\mathbf{x}_1, \mathbf{x}_2, t)dA_1dA_2$ is the probability of finding a pair of brownian bugs with one member in the dA_1 surrounding $\mathbf{x}_1$ and the other member in the dA_2 surrounding $\mathbf{x}_2$. For example, a Poisson point process with uniform concentration C_0 has $G = C_0^2$; departures of G from C_0^2 indicate pair correlations. In particular, because births occur at the same location as the parent, births increase the correlation G above C_0^2 for small $\mathbf{x}_1 - \mathbf{x}_2$. Because the pair correlation function G is the Fourier transform of the 'coarse-grained' concentration spectrum²⁴, we can relate this analysis to spectral descriptions of plankton patchiness^{23,9}.

For an isotropic and spatially homogeneous process, such as those in Figs 1 and 2, G depends only on the pair separation $r \equiv |\mathbf{x}_1 - \mathbf{x}_2|$. Then it is convenient to define the radial density function $g(r, t)$ by $G(\mathbf{x}_1, \mathbf{x}_2, t) = C_0^2 g(r, t)$ where C_0 is the uniform concentration. Because the pair correlations disappear at great distances, $g \rightarrow 1$ as $r \rightarrow \infty$. In a two-dimensional space, the expected number of brownian bugs located in the annulus $(r, r + dr)$ surrounding a typical brownian bug is $C_0 g(r, t) 2\pi r dr$; thus, by making a histogram of pair separations, we can estimate g from the results of a simulation (Fig. 3).

The solid curve in Fig. 3 shows that without random advection ($U = 0$) there is a greatly enhanced probability of finding one brownian bug in the close neighbourhood of another. The effect of random advection is to reduce g at $r = 0$ and to produce a long tail signifying slowly decaying pair correlations at large r (Fig. 3). The reduction at $r = 0$ quantifies the rapid advective separation of nearly coincident pairs, whereas the tail is the signature of the correlations that exist along the filaments in Fig. 2b. At moderate values of the stretching parameter U in equation (3), this tail is well described by $g \approx r^{-2}$ (Fig. 3): advection reduces reproductive pair correlations at small r , but enhances these correlations at large r .

The pair correlation function G is useful as a quantitative diagnostic of patchiness. But theoretical progress is also possible because for the brownian bug process $G(r, t)$ satisfies

$$G_t = 2Dr^{1-d}(r^{d-1}G_r)_r + 2(\lambda - \mu)G + \gamma r^{1-d}(r^{d+1}G_r)_r + 2\lambda C\delta(\mathbf{r}) \quad (2)$$

In this equation, d is the dimension of the space ($d = 2$ in all our simulations). The first term on the right-hand side is the diffusive separation of pairs with a diffusivity $2D$ because each member of the pair is an independent random walker. The second term on the right is the production of new pairs, which occurs if the birth rate λ exceeds the death rate μ . The third term on the right describes the separation of pairs by rapidly decorrelating random advection²¹, this particular form applies provided that r is much less than the smallest scale of the velocity field. The final term is production of coincident pairs (at $r = 0$) by birth.

If $\lambda = \mu$ then we can obtain a steady ($G_t = 0$) solution of equation (2). These are 'constant-flux' solutions: the reproductive source produces pairs at the origin ($r = 0$) of the pair-separation space. Then a combination of brownian motion and advective stretching transports these pairs outwards to larger values of r . Where r is much greater than $\sqrt{D/\gamma}$, advective stretching is the dominant process and the steady solution has $G \propto r^{-d}$ (this is the r^{-2} regime in Fig. 3). These steady analytic solutions of equation (2) are shown as dash-dot curves in Fig. 3. The large differences between theory and simulation at small r signifies the failure of the continuum approximation equation (2) once r is comparable to the step length of the random walk.

The brownian bug process is a simple model of plankton patchiness with two ingredients the production of concentration fluctuations by reproduction and the reduction in length scale of these fluctuations by stirring. It is crucial that the reproductive source term, $\delta(\mathbf{r})$ in equation (2), forces all wavenumbers equally. (We note

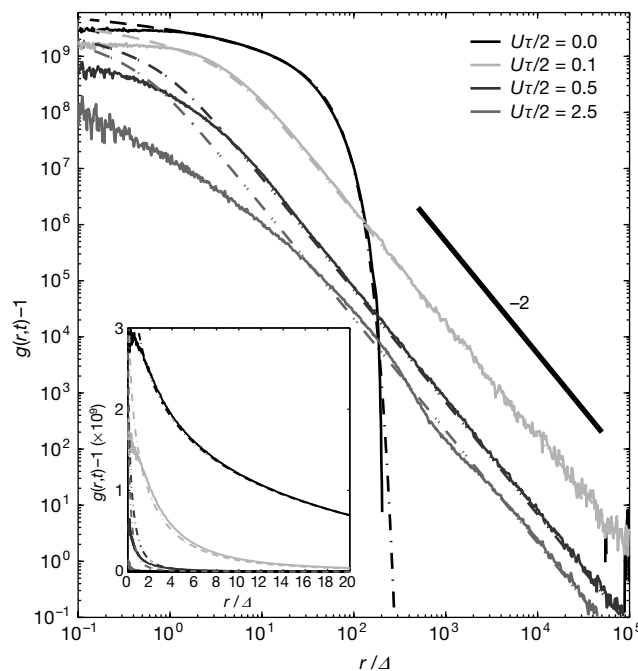


Figure 3 Logarithmic and linear plots of $g(r, t) - 1$ versus r/Δ . $t = 1,000\tau$, $N_0 = 20,000$, $\Delta = 10^{-7}$ and $U\tau/2 = 0, 0.1, 0.5, 2.5$. The dash-dot curves are

analytic solutions of equation (2). Inset, linear plot. The thick line labelled -2 indicates the r^{-2} scaling predicted by equation (2).

that if one Fourier transform acts on equation (2), then $\delta(\mathbf{r}) \rightarrow 1$, showing that the final term forces a white spectrum.) Other models of plankton patchiness^{2,3,9,11} follow Batchelor²⁰ in assuming that concentration variations are forced externally only at small wavenumbers. The brownian bug process shows that the fluctuations can be generated by reproduction and that this process pumps variance into both large and small spatial scales. The common factor is that stirring reduces the length scale of concentration fluctuations so that diffusion can become effective. □

Methods

Conceptually the brownian bug model is a population of random-walking individuals evolving in continuous time and space and simultaneously undergoing a branching process²⁵. In probability theory this is known as a superbrownian process²⁶. To simulate this on a computer we must discretize.

The initial condition ($t = 0$) is prepared by randomly and independently placing $N_0 \gg 1$ brownian bugs (idealized as points) in the $L \times L$ square (periodically extended in order to avoid dealing with reflection boundary conditions). For visualization each brownian bug is tagged with a colour that varies smoothly with the y coordinate. The simulation is advanced through time in increments of a 'cycle' of duration τ . Each cycle consists of three steps: (1) random birth and death; (2) brownian motion; and (3) advective stirring. In step (1) each brownian bug reproduces by binary fission (with probability p) or dies (with probability q) or remains unchanged (with probability $1 - p - q$). When a lucky brownian bug divides, the offspring is placed on top of the parent and parents transmit their colour to their offspring. In step (2), bug k is displaced to a new position $\mathbf{x}'_k(t) = \mathbf{x}_k(t) + \delta\mathbf{x}_k(t)$. The components of $\delta\mathbf{x}_k$ are independent and identically distributed gaussian random variables, each with r.m.s. value Δ (that is, brownian motion with the diffusivity $D = \Delta^2/2\tau$). These independent displacements separate coincident parent-offspring pairs. In step (3) we use the random map procedure²⁴

$$\begin{aligned}x_k(t + \tau) &= x'_k(t) + (U\tau/2) \cos[ky'_k(t) + \varphi(t)] \\y_k(t + \tau) &= y'_k(t) + (U\tau/2) \cos[kx'_k(t) + \theta(t)]\end{aligned}$$

where $\varphi(t)$ and $\theta(t)$ are independent random phases and $k = 2\pi/L$.

In principle, we can approach the continuous limit in equations (1) and (2) by taking $\tau \rightarrow 0$ while holding the parameters $\lambda = p/\tau$, $\mu = q/\tau$, $D = \Delta^2/2\tau$ and $(kU)^2\tau$ fixed.

Step (1) is a Galton-Watson branching process²⁵ and because the probabilities of birth and death are equal (we take $p = q = 1/2$) we are at the critical point. In this case, even though the average population is always N_0 , the actual population has large fluctuations and total extinction is certain if the simulation lasts too long. We avoid this issue by making N_0 much larger than the number of generations. The suppression of extinction is a consequence of approaching the 'thermodynamic limit' ($N_0 \rightarrow \infty$ with $C_0 = N_0/L^2$ fixed).

The stretching parameter γ in equation (2) was determined for each value of $U\tau$ as the

best-fit slope of $(1/d)\langle \ln r(t) \rangle$ against $t \cdot \langle \ln r(t) \rangle$ is an average obtained from an ensemble of 800 particle pairs (without diffusion and birth/death) initially separated by $r(0) = 10^{-7}$. The growth rate λ was obtained by considering the case $U = 0$ and $\lambda = \mu$ and then least-squares-fitting the similarity solution of equation (2) to the results of a simulation in the range Δ to 200Δ . The resulting estimate of λ conforms with the expected result that $\lambda \approx 1/2\tau$.

Received 14 November 2000; accepted 4 June 2001.

1. Flierl, G., Grünbaum, D., Levin, S. & Olson, D. From individuals to aggregations: the interplay between behaviour and physics. *J. Theor. Biol.* **196**, 397–454 (1999).
2. Abraham, E. R. The generation of plankton patchiness by turbulent stirring. *Nature* **391**, 577–580 (1998).
3. Denman, K. L., Okubo, A. & Platt, T. The chlorophyll fluctuation spectrum in the sea. *Limnol. Oceanogr.* **22**, 1033–1038 (1977).
4. Franks, P. J. S. Spatial pattern in dense algal blooms. *Limnol. Oceanogr.* **42**, 1297–1305 (1997).
5. Kierstead, H. & Slobodkin, L. B. The size of water masses containing plankton blooms. *J. Mar. Res.* **12**, 141–147 (1953).
6. Neufeld, Z., López, C. & Haynes, P. H. Smooth-filament transition of active tracer fields stirred by chaotic advection. *Phys. Rev. Lett.* **82**, 2606–2609 (1999).
7. Robinson, A. R. On the theory of advective effects on biological dynamics in the sea. *Proc. R. Soc. Lond. A* **453**, 2295–2324 (1997).
8. Steele, J. H. & Henderson, E. W. A simple model for plankton patchiness. *J. Plankton Res.* **14**, 1397–1403 (1992).
9. Gower, J. F. R., Denman, K. L. & Holyer, R. J. Phytoplankton patchiness indicates the fluctuation spectrum of mesoscale oceanic structure. *Nature* **288**, 157–159 (1980).
10. Lesieur, M. & Sadourny, R. Satellite-sensed turbulent ocean structure. *Nature* **294**, 673 (1980).
11. Abraham, E. R. *et al.* Importance of stirring in the development of an iron-fertilized phytoplankton bloom. *Nature* **407**, 727–730 (2000).
12. Durrett, R. & Levin, S. The importance of being discrete (and spatial). *Theor. Pop. Biol.* **46**, 363–394 (1994).
13. Levin, S. & Segel, L. A. Hypothesis for origin of planktonic patchiness. *Nature* **259**, 659 (1976).
14. Segel, L. A. & Jackson, J. L. Dissipative structure: an explanation and an ecological example. *J. Theor. Biol.* **37**, 545–559 (1972).
15. Skellam, J. G. Random dispersion in theoretical populations. *Biometrika* **38**, 196–218 (1951).
16. Keeling, M. J., Mezić, L., Hendry, R. J., McGlade, J. & Rand, D. A. Characteristic length scales of spatial models in ecology via fluctuation analysis. *Phil. Trans. R. Soc. Lond. B* **352**, 1589–1601 (1997).
17. Pascual, M. & Levin, S. A. From individuals to population densities: searching for the intermediate scale of nontrivial determinism. *Ecology* **80**, 2225–2236 (1999).
18. Rand, D. A. & Wilson, H. B. Using spatio-temporal chaos and intermediate scale determinism in artificial ecologies to quantify spatially extended systems. *Proc. R. Soc. Lond. B* **259**, 55–63 (1995).
19. Batchelor, G. K. The effect of turbulence on material lines and surfaces. *Proc. R. Soc. Lond. B* **213**, 349–366 (1952).
20. Batchelor, G. K. Small-scale variation of convected quantities like temperature in turbulent fluid. Part 1. General discussion and the case of small conductivity. *J. Fluid Mech.* **5**, 113–133 (1959).
21. Kraichnan, R. H. Convection of a passive scalar by a quasi-uniform random stretching field. *J. Fluid Mech.* **64**, 737–762 (1974).
22. Pierrehumbert, R. T. On tracer microstructure in the large-eddy dominated regime. *Chaos Solitons Fractals* **4**, 1091–1110 (1994).

23. van Kampen, N. G. *Stochastic Processes in Physics and Chemistry* Ch. 2 (North-Holland, Amsterdam, 1981).
24. Pierrehumbert, R. T. in *Nonlinear Phenomena in Atmospheric and Oceanic Sciences* Ch. 2 (Springer, New York, 1991).
25. Harris, T. E. *The Theory of Branching Processes* Ch. 1 (Springer, Berlin, 1963).
26. Le Gall, J. F. *Spatial Branching Processes, Random Snakes and Partial Differential Equations*, Lectures in Mathematics, ETH Zürich (Birkhäuser, Basel, 1999).

Acknowledgements

We thank E. Ben-Naim, R. Durrett, G. Flierl, P. Krapivsky, P. Morrison and F. Williams for discussion.

Correspondence and requests for materials should be addressed to W.R.Y. (e-mail: wryoung@uconn.edu).

Evolution of digital organisms at high mutation rates leads to survival of the flattest

Claus O. Wilke*, Jia Lan Wang*, Charles Ofria†, Richard E. Lenski† & Christoph Adami*‡

* Digital Life Laboratory, Mail Code 136-93, California Institute of Technology, Pasadena, California 91125, USA

† Center for Biological Modeling, Michigan State University, East Lansing, Michigan 48824, USA

‡ Jet Propulsion Laboratory, Mail Code 126-347, California Institute of Technology, Pasadena, California 91109, USA

Darwinian evolution favours genotypes with high replication rates, a process called 'survival of the fittest'. However, knowing the replication rate of each individual genotype may not suffice to predict the eventual survivor, even in an asexual population. According to quasi-species theory, selection favours the cloud of genotypes, interconnected by mutation, whose average replication rate is highest^{1–5}. Here we confirm this prediction using digital organisms that self-replicate, mutate and evolve^{6–9}. Forty pairs of populations were derived from 40 different ancestors in identical selective environments, except that one of each pair experienced a 4-fold higher mutation rate. In 12 cases, the dominant genotype that evolved at the lower mutation rate achieved a replication rate >1.5-fold faster than its counterpart. We allowed each of these disparate pairs to compete across a range of mutation rates. In each case, as mutation rate was increased, the outcome of competition switched to favour the genotype with the lower replication rate. These genotypes, although they occupied lower fitness peaks, were located in flatter regions of the fitness surface and were therefore more robust with respect to mutations.

Mutation and natural selection are the two most basic processes of evolution, yet the study of their interplay remains a challenging area for theoretical and empirical research. Recent studies have examined the effect of mutation rate on the speed of adaptive evolution^{10,11} and the role of selection in determining the mutation rate itself^{12–14}. Quasi-species models predict a particularly subtle interaction: mutation acts as a selective agent to shape the entire genome so that it is robust with respect to mutation^{1–5}. (See refs 15–18 for related predictions expressed in other terms.) In particular, selection in an asexual population should maximize the overall replication rate of a cloud of genotypes connected by mutation, rather than fix any one genotype that has the highest replication rate. Thus, a fast-replicating organism that occupies a high and narrow peak in the fitness landscape—where most nearby mutants

are unfit—can be displaced by an organism that occupies a lower but flatter peak. Thus, 'survival of the flattest' may be as important as 'survival of the fittest' at high mutation rates. This prediction has proved difficult to test experimentally, but a recent study¹⁹ with an RNA virus reported that two populations, derived from a common ancestor, have mutational neighbourhoods with different distributions of fitness effects.

Direct evidence for the displacement of a fast replicator by a more robust, slower one must come from experiments in which such organisms are squarely pitted against each other. The systematic (repeatable) winner of such a competition is, in effect, the fitter one, although the loser may have the higher replication rate. For example, imagine that a particular mutation yields a more robust genotype, but at the cost of a slightly lower replication rate. It is an empirical question whether the advantage of the mutational robustness is sufficient to offset its disadvantage in terms of replication rate. Quasi-species theory predicts that, under appropriate conditions (high mutation pressure), such a mutation can be fixed in an evolving population, despite its lower replication rate. This prediction does not depend on the details of the organism chosen for experiments, but only on mutation rate, replication speed, and robustness to mutations. Microorganisms, such as bacteria and viruses, are often used to test evolutionary theories, and competition experiments are typically performed to quantify fitness in the course of these tests. However, it would be difficult to disentangle the contributions of replication rate and robustness, because competitions measure the combined effect of both processes. Here, we use a more convenient system for disentangling these effects: digital organisms that live in, and adapt to, a virtual world created for them inside a computer.

Digital organisms are self-replicating computer programs that compete with one another for CPU (central processing unit) cycles, which are their limiting resource. Digital organisms have genomes (series of instructions) and phenotypes that are obtained by the execution of their genomic programs. The evolution of these programs is not simulated in the conventional (numerical) sense. Instead, they physically inhabit a reserved space in the computer's memory (an 'artificial Petri dish'), and they must copy their own genomes. Moreover, their evolution does not proceed towards a target specified in advance; instead it proceeds in an open-ended manner to produce phenotypes that are more successful in a particular environment. Digital organisms acquire resources (CPU cycles) by performing certain logical functions, such as biochemical organisms catalyse exothermic reactions to obtain energy. They lend themselves to evolutionary experiments because their environment can be readily manipulated to examine the importance of various selective pressures. The only environmental

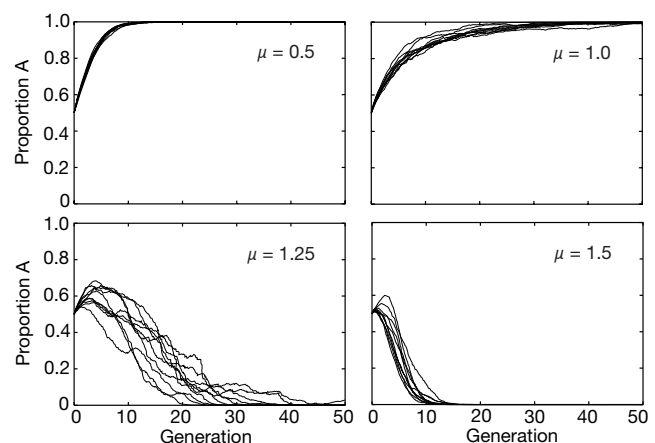


Figure 1 Competitions for one pair of organisms at four different mutation rates. Organism A replicates 1.96 times faster than B. μ , Genomic mutation rate.

factor varied here is mutation rate (such as might be achieved by varying temperature or radiation with biochemical organisms). Digital organisms replicate by copying their genomes one instruction at a time, just as biochemical organisms copy DNA (or RNA) base by base. This process yields overall copy fidelity $F \equiv e^{-\mu}$, where $\mu = RL$ is the mutation rate per genome, R is the error rate per instruction copied, and L is the length of the genomic sequence.

Populations of digital organisms were used to test directly the prediction from quasi-species theory that natural selection can favour genotypes with slower replication, provided they occupy flatter peaks surrounded by mutants that are also reasonably fit. We evolved 40 pairs of digital organisms; one of each pair adapted to a low (A) mutation rate and the other to a high (B) rate. Evolution at high mutation rate creates a selective force that favours robustness^{20,21} and thus increases the likelihood of obtaining genotypes for competition experiments that are informative in the present context. Among all 40 pairs, we found 12 in which A achieved a more than 1.5-fold advantage in replication relative to B. For these 12 pairs, A and B were then mixed in equal proportions and allowed to compete across a range of mutation rates. After 50 generations, we measured the percentage of organisms derived from A in the mixed population.

For all 12 pairs, A excluded B at low mutation rates, whereas B prevailed at high mutation rates. Figure 1 shows the actual dynamics of competition for one pair. Figure 2 shows, for this same pair, that the competitive reversal reflects a shift toward less fit genotypes, which is more pronounced for A than B. These data therefore demonstrate that A occupied a higher but narrower fitness peak, whereas B was on a lower but broader peak. We note that for all pairs and all mutation rates, A could increase to its carrying capacity in the absence of its competitor. Thus, the reason for the extinction of A was not simply an unbearably high mutation rate.

We then sought an effective measure of the breadth of fitness

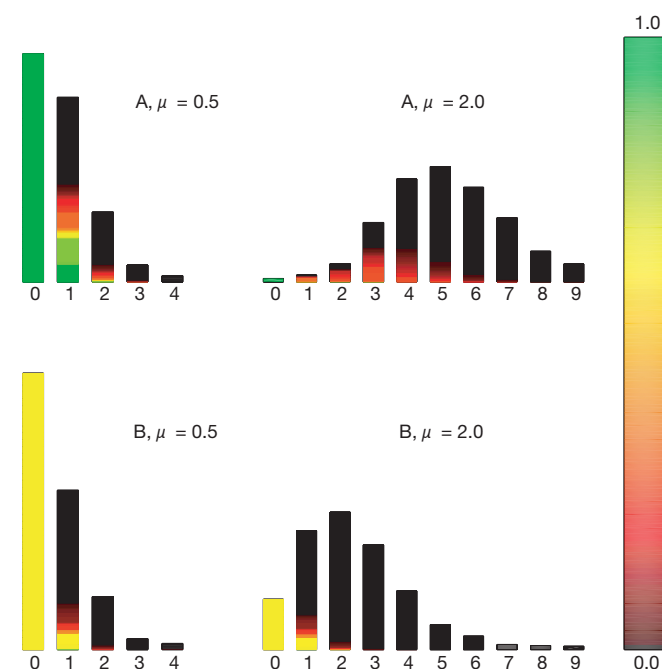


Figure 2 Genotype distributions after 15 generations from populations seeded uniformly with either A or B. Numbers below the bars are Hamming distances from the seeding genotype; bar heights are relative frequencies of genotypes at those distances. Colour coding indicates the intrinsic replication rates of all the genotypes. The colour scale was normalized to the intrinsic replication rate of A. The pair of organisms in this graph is the same pair for which the dynamics of competition were shown in Fig. 1. The right tails of the distributions are truncated for the purposes of illustration only.

peak. Such a measure would allow us to define the critical mutation rate, μ_{crit} , at which A and B perform equally well; and thereby predict, for any given mutation rate, whether A or B would prevail in competition. The digital organisms in our study are asexual, that is, there is no recombination between individuals associated with their replication. We initially tried an approach that relies on an approximation from population genetics theory, which says that the equilibrium genetic load (reduction in mean fitness) from deleterious mutations in an asexual population is equal to the genomic rate of deleterious mutation^{12–24}. For each organism in the 12 pairs, we examined every possible one-step mutation⁸ to determine the exact proportion of deleterious mutations, which could be scaled by the mutation rate to predict μ_{crit} . To our surprise, this approach was highly unsatisfactory for predicting the outcome of competition or the critical mutation rate at which the competitive reversal occurs (data not shown). The weakness of this approach lies in the implicit assumption that competing populations remain tightly centred on the original genotypes. This assumption is clearly false, as seen in Fig. 2 by the large proportion of genotypes that differ by two or more steps from the original type.

We thus characterized each organism's mutational neighbourhood in a different way. Specifically, we measured the actual growth rate of the population spawned from a single organism across a range of mutation rates. The decay in growth rate with increasing mutation rate was well described by:

$$w(\mu) = w_0 \exp(-a\mu - b\mu^2), \quad (1)$$

where μ is the genomic mutation rate, w_0 is the intrinsic replication rate of the organism, and a and b are parameters estimated by fitting the data. (A similar function can be derived from quasi-species theory, see Supplementary Information.) Because the function $w(\mu)$ describes the realized growth rate of a population at a given mutation rate, the outcome of a competition between A and B should be determined by whose realized growth rate $w(\mu)$ is greater at a certain mutation rate, not by whose intrinsic replication rate w_0 is larger. The predicted critical mutation rate, μ_{crit} , is then simply the value where the two organisms' realized growth rates cross. (We note that the function $w(\mu)$ used here is fundamentally different from the decay function in ref. 8. Our function describes the decay of the realized growth rate of a population with increasing mutation rate, whereas the earlier function reflects the decay of intrinsic

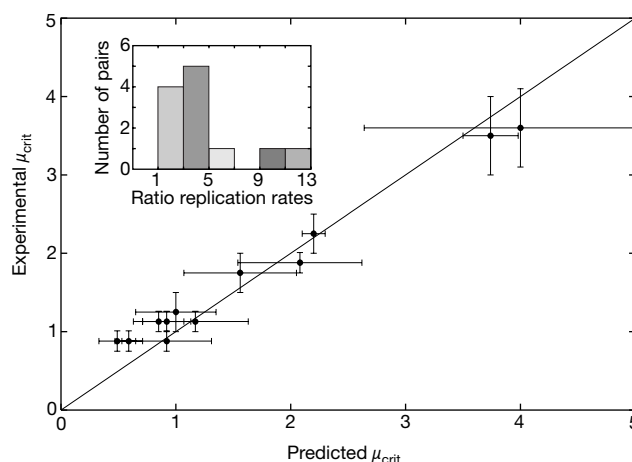


Figure 3 Critical mutation rates obtained from competition experiments versus predicted estimates based on each organism's intrinsic replication rate, w_0 , and mutational robustness parameters, a and b . Points along the diagonal line would imply perfect agreement between experiment and prediction; the actual correlation coefficient is 0.991 ($P < 0.0001$). The inset shows the distribution of intrinsic replication rate ratios ($w_{0,A}/w_{0,B}$) for the 12 pairs.

replication rate as a function of genetic distance averaged over all mutants.) To test the utility of our new measure of peak breadth, we obtained the function $w(\mu)$ for all 24 organisms; for each corresponding pair we then calculated the predicted μ_{crit} value. Figure 3 shows the experimental μ_{crit} values (obtained by interpolation from the competitions) plotted against the predicted values. In all 12 cases, the experimental and predicted values agree well, and the overall correlation coefficient is 0.991 ($P < 0.0001$).

Thus we have demonstrated that faster replicating organisms can easily be out-competed at high mutation rates by organisms that replicate more slowly, if the latter obtain sufficient support from their mutational neighbourhood. Even a 12-fold difference in replication rate could be overcome by greater mutational robustness of the slower replicator (inset to Fig. 3). We emphasize that this robustness was not caused by any difference in replication fidelity, but rather by differences in 'canalization' with respect to mutational perturbation¹⁷. We also showed a widespread trade-off between intrinsic replication rate and mutational robustness, which arose during divergence from a common ancestor in environments that differed only in the imposed mutation rate. These findings demonstrate the importance of the mutational cloud, as described by the quasi-species model. The mutation rates where we saw these effects were on the order of one per genome per generation. Such mutation rates are not unusually high; they occur in RNA viruses^{25,26} and many DNA-based eukaryotes^{26–28}.

One difference between the digital organisms studied here and biochemical organisms is that the latter can control, to some degree, their own mutation rates through DNA editing and repair. Thus, mutation rate may depend on the particular genotype as well as the environment. Nonetheless, evolution and competition experiments with bacteria or viruses could be performed at different mutation rates by varying the concentration of some mutagenic agent. We expect that more robust organisms would prevail over faster replicating, but more brittle, organisms at high mutation rates. □

Methods

All experiments were performed using the Avida platform, version 1.4, which can be obtained, along with the configuration files necessary to reproduce our experiments, from <http://dallab.caltech.edu/pubs/nature01/>. The particular organisms used here were taken from a pool generated in another study⁸. The organisms had genome lengths between 54 and 314 instructions, and they typically performed between 20 and 30 one-, two-, or three-input logical operations. In all experiments reported here, we disabled changes in genome length, which ensured that the genome-wide mutation rate did not change during evolution of the pairs from a common ancestor. We also prevented the organisms from evolving new computational functions by rewarding only logical operations that their common ancestor already possessed.

Adaptation

For the adaptation phase, we used 40 previously diverged ancestors to seed 40 pairs of populations at low and high mutation rates of 0.5 and 2.0 per genome per generation, respectively. Apart from the mutation rate, the selective environment was identical for each pair derived from the same ancestor. After 1,000 generations, we extracted the most abundant genotype from each population, giving us 40 pairs of organisms in which one member, designated A, was adapted to the lower mutation rate and the other, designated B, was adapted to the higher mutation rate.

Among the 40 pairs, we found one pair where both dominant organisms were nonviable (this can occur, for example, if another abundant genotype has a genome for which a lethal miscopy occurs with a high probability). In three cases, B evolved a higher replication rate than A. Because we were interested in situations where B could out-compete A despite having a lower replication rate, we ignored these cases. Among the 36 remaining pairs where A had a higher replication rate than B, we found 24 pairs with replication-rate ratios between 1.0 and 1.5, and 12 pairs with ratios above 1.5. We used the latter group for competition experiments, as these should both provide a clearer signal of the phenomenon of interest and provide a more stringent hurdle to be overcome by differences in mutational robustness.

Competition

The competition experiments lasted 50 generations. We seeded the mixed population with 50% each of A and B. We marked A with an inherited label, so that we could measure the percentage of descendants of each type in the mixed population. Population size was fixed at 3,600. For each pair of A and B, we ran competitions at genomic mutation rates of 0.5, 1.0, 1.5, 2.0, 2.5 and 3.0, with tenfold replication. We also ran additional competition at

intermediate rates to locate the critical mutation rate more closely, and we tested up to a mutation rate of 4.0 in two cases where A prevailed at mutation rates of 3.0 and below.

To obtain an independent measure of the mutational robustness of each organism, we seeded a population with that single genotype and measured mean fitness (number of organisms born per unit time) after 15 generations, at genomic mutation rates ranging from 0 to 3.0 in steps of 0.5. The mean fitness at vanishing mutation rate corresponds to the intrinsic replication rate w_0 of that genotype. The robustness parameters, a and b , were determined from a nonlinear fit of equation (1) to the data.

Critical mutation rate

We determined the critical mutation rate from the competition experiments as follows. For all 12 pairs, the population consisted almost entirely of the descendants of only a single competitor (that is, one of the two had disappeared) after 50 generations, except at mutation rates near the critical point, where more time was needed for extinction. The critical mutation rate is the midpoint between the highest rate where A prevailed and the lowest rate where B prevailed. The error is one-half the corresponding interval.

The independently predicted critical mutation rate is given by the smallest positive root to the quadratic equation $(a_A - a_B)\mu + (b_A - b_B)\mu^2 = \ln(w_{0,A}/w_{0,B})$, where subscripts indicate organism A or B, a and b are robustness parameters from equation (1), and w_0 is the intrinsic replication rate. The corresponding error is obtained by propagating the errors of the robustness parameters from the fit of equation (1).

Received 13 December 2000; accepted 7 June 2001.

1. Eigen, M. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465–523 (1971).
2. Eigen, M. & Schuster, P. *The Hypercycle: A Principle of Natural Self-Organization* (Springer, Berlin, 1979).
3. Schuster, P. & Swetina, J. Stationary mutant distributions and evolutionary optimization. *Bull. Math. Biol.* **50**, 636–660 (1988).
4. Eigen, M., McCaskill, J. & Schuster, P. The molecular quasi-species. *Adv. Chem. Phys.* **75**, 149–263 (1989).
5. Nowak, M. A. What is a quasi-species? *Trends Ecol. Evol.* **7**, 118–121 (1992).
6. Ray, T. S. in *Artificial Life II* (eds Langton, C. G., Taylor, C., Farmer, J. D. & Rasmussen, S.) 372–408 (Addison-Wesley, Redwood City, 1991).
7. Adami, C. *Introduction to Artificial Life* (Springer, New York, 1998).
8. Lenski, R. E., Ofria, C., Collier, T. C. & Adami, C. Genome complexity, robustness and genetic interactions in digital organisms. *Nature* **400**, 661–664 (1999).
9. Adami, C., Ofria, C. & Collier, T. C. Evolution of biological complexity. *Proc. Natl Acad. Sci. USA* **97**, 4463–4468 (2000).
10. Gerrish, P. J. & Lenski, R. E. The fate of competing beneficial mutations in an asexual population. *Genetica* **102/103**, 127–144 (1998).
11. De Visser, J. A. G. M., Zeyl, C. W., Gerrish, P. J., Blanchard, J. L. & Lenski, R. E. Diminishing returns from mutation supply rate in asexual populations. *Science* **283**, 404–406 (1999).
12. Taddei, F. *et al.* Role of mutator alleles in adaptive evolution. *Nature* **387**, 700–702 (1997).
13. Sniegowski, P. D., Gerrish, P. J. & Lenski, R. E. Evolution of high mutation rates in experimental populations. *E. coli. Nature* **387**, 703–705 (1997).
14. Johnson, T. The approach to mutation–selection balance in an infinite asexual population, and the evolution of mutation rates. *Proc. R. Soc. Lond. B* **266**, 2389–2397 (1999).
15. Waddington, C. H. Canalization of development and genetic assimilation of acquired characters. *Nature* **183**, 1654–1655 (1959).
16. Nowak, M. A., Boerlijst, M. C., Cooke, J. & Maynard Smith, J. Evolution of genetic redundancy. *Nature* **388**, 167–171 (1997).
17. Wagner, G. P., Booth, G. & Bagheri-Chaichian, H. A population genetic theory of canalization. *Evolution* **51**, 329–347 (1997).
18. Freeland, S. J. & Hurst, L. D. Load minimization of the genetic code: history does not explain the pattern. *Proc. R. Soc. Lond. B* **265**, 2111–2119 (1998).
19. Burch, C. L. & Chao, L. Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature* **406**, 625–628 (2000).
20. van Nimwegen, E., Crutchfield, J. P. & Huynen, M. Neutral evolution of mutational robustness. *Proc. Natl Acad. Sci. USA* **96**, 9716–9720 (1999).
21. Wike, C. O. Adaptive evolution on neutral networks. *Bull. Math. Biol.* (in the press).
22. Haldane, J. B. S. The effect of variation on fitness. *Am. Nat.* **71**, 337–349 (1937).
23. Muller, H. J. Our load of mutations. *Am. J. Hum. Genet.* **2**, 111–176 (1950).
24. Charlesworth, B. Mutation–selection balance and the evolutionary advantage of sex and recombination. *Genet. Res.* **55**, 199–221 (1990).
25. Drake, J. W. Rates of spontaneous mutation among RNA viruses. *Proc. Natl Acad. Sci. USA* **90**, 4171–4175 (1993).
26. Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. W. Rates of spontaneous mutation. *Genetics* **148**, 1667–1686 (1998).
27. Eyre-Walker, A. & Keightley, P. D. High genomic deleterious mutation rates in hominids. *Nature* **397**, 344–347 (1999).
28. Lynch, M. *et al.* Spontaneous deleterious mutation. *Evolution* **53**, 645–663 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work was supported by the National Science Foundation. Part of this work as carried out at the Jet Propulsion Laboratory, under a contract with the National Aeronautics and Space Administration. J.L.W. acknowledges the support of an Axline SURF fellowship.

Correspondence and requests for materials should be addressed to C.O.W. (e-mail: wike@caltech.edu).

Evolution and transmission of stable CTL escape mutations in HIV infection

Philip J. R. Goulder^{*,†}, Christian Brander^{*}, Yanhua Tang^{*}, Cecile Tremblay^{*}, Robert A. Colbert[‡], Marylyn M. Addo^{*}, Eric S. Rosenberg^{*}, Thi Nguyen^{*}, Rachel Allen[§], Alicja Trocha^{*}, Marcus Altfeld^{*}, Suqin He^{*}, Michael Buncell^{||}, Robert Funkhouser[¶], Stephen I. Pelton[#], Sandra K. Burchett[†], Kenneth McIntosh[†], Bette T. M. Korber[¶] & Bruce D. Walker^{*}

^{*} Partners AIDS Research Center, Massachusetts General Hospital and Division of AIDS, Harvard Medical School, Boston, Massachusetts 02114, USA

[†] Division of Infectious Diseases, The Children's Hospital, and Division of AIDS, Harvard Medical School, Boston, Massachusetts 02115, USA

[‡] William S. Rowe Division of Rheumatology, Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, Ohio 45229-3039, USA

[§] Department of Pathology, Cambridge University, Cambridge CB2 1QP, UK

^{||} Oxford Transplant Centre, Churchill Hospital, Oxford OX3 7LJ, UK

[¶] Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

[#] Section of Pediatric Infectious Diseases, Boston Medical Center, Boston, Massachusetts 02118, USA

Increasing evidence indicates that potent anti-HIV-1 activity is mediated by cytotoxic T lymphocytes (CTLs)^{1–3}; however, the effects of this immune pressure on viral transmission and evolution have not been determined. Here we investigate mother–child transmission in the setting of human leukocyte antigen (HLA)-B27 expression, selected for analysis because it is associated with prolonged immune containment in adult infection⁴. In adults, mutations in a dominant and highly conserved B27-restricted Gag CTL epitope lead to loss of recognition and disease progression^{4–6}. In mothers expressing HLA-B27 who transmit HIV-1 perinatally, we document transmission of viruses encoding CTL escape variants in this dominant Gag epitope that no longer bind to B27. Their infected infants target an otherwise subdominant B27-restricted epitope and fail to contain HIV replication. These CTL escape variants remain stable without reversion in the absence of the evolutionary pressure that originally selected the mutation. These data suggest that CTL escape mutations in epitopes associated with suppression of viraemia will accumulate as the epidemic progresses, and therefore have important implications for vaccine design.

More than 85% of HLA-B27-positive adults infected with HIV-1 target the p24 Gag epitope KK10 (KRWILGLNK, residues 131–140)^{4–7}, a strongly conserved epitope⁸ associated with good clinical outcome^{4,9}. In contrast, only two out of six B27-positive children examined targeted this CTL epitope ($P = 0.02$, Fisher's exact test). Mothers of four of these children were available for HLA typing. All three children who shared HLA-B27 with their mothers (048-C, 049-C and 002-C) failed to target the KK10 epitope (Fig. 1, $P = 0.01$, Fisher's exact test). In the one child positive for B27 whose mother did not express B27 (child 043-C), a dominant response to KK10 was detected (Fig. 1). This child resembles adults with HLA-B27, as at 7.8 years of age he is a long-term non-progressor, with a CD4 count of greater than 800 cells mm⁻³ and an undetectable viral load since the first measurement at age 4 (less than 50 RNA copies ml⁻¹ plasma) on 2',3'-dideoxyinosine/3'-azido-2'-deoxythymidine (ddI/AZT) therapy. This is exceptional by paediatric standards¹⁰, with parameters such as these being seen in only 5 out of 130 infected children currently undergoing treatment at outpatient clinics of The Children's Hospital in Boston.

To investigate the absence of detectable CTLs to the dominant HLA-B27-restricted KK10 epitope in these children, virus was analysed by population and clonal sequencing in the four

mother–infant pairs for whom samples were available (Table 1). Sequences were also confirmed by analysis of cultured virus (not shown). All three mothers expressing HLA-B27 had KK10 mutations identical to those detected in their children; these mutations included the HLA-B27 anchor position-2 (P2) at which arginine is required for peptide binding¹¹. These P2 variants (arginine to threonine at position 132, R132T) failed to bind B27 (Fig. 2a), and no CD8 T-cell response was detectable either to KK10 or to the autologous variant in these children and mothers. In addition to the P2 mutation, the presumed compensatory P6 mutation (L136M) was also detected⁵. In the one mother (043-M) who did not express HLA-B27, the autologous virus was fully wild type within the Gag epitope, as was the P2 anchor position in her child (Table 1). This is consistent with the hypothesis that the virus transmitted from mother to child had not been under immune selection pressure in this region, as the child inherited HLA-B27 from his father (043-D). CTL clones from this child readily inhibited autologous virus replication *in vitro*, but were unable to inhibit the P2 variant isolated from one of the children with a mutated KK10 epitope (Fig. 2b).

To address whether these mutations arose independently, we determined the frequency of P2 anchor-position changes in the KK10 epitope. In only 2 out of 14 cases evaluated (mothers 002-M

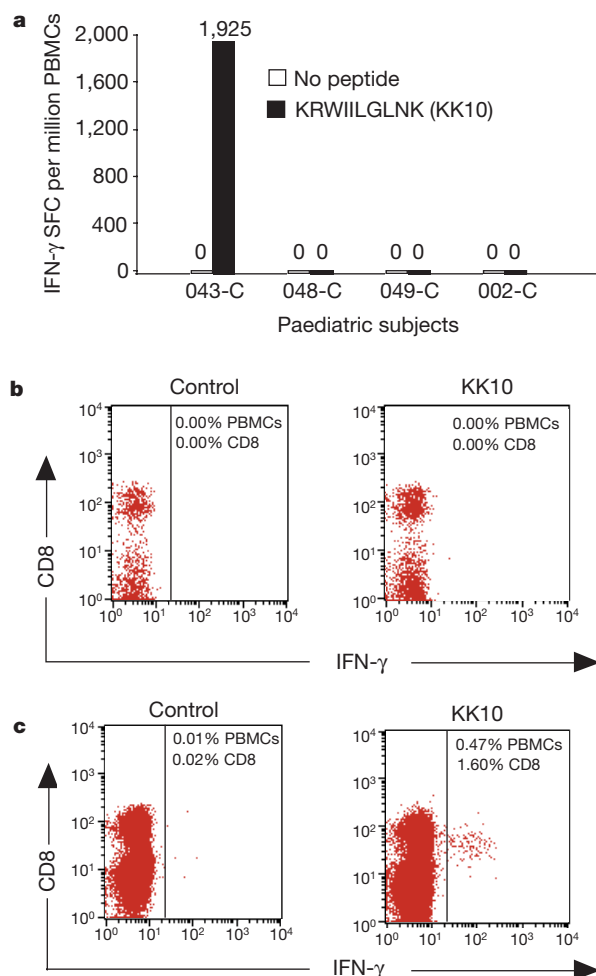


Figure 1 Recognition of the B27-restricted Gag epitope KRWILGLNK in HLA-B27-positive, HIV-infected children. **a**, Recognition of the B27-restricted epitope KRWILGLNK (KK10) in ELISPOT assays by HLA-B27-positive children. All three children with no response to KK10 had mothers who expressed HLA-B27. **b**, Intracellular IFN- γ staining assays using cells from 048-C stimulated with no peptide (left panel) and KK10 wild-type peptide (10 μ M, right panel). **c**, Intracellular IFN- γ staining assays using cells from 043-C (peptide stimulations as in **b**).

and 048-M) were there Arg→Thr variants observed at this critical anchor position. Thus, even assuming that all perinatally infected children with HLA-B27 inevitably develop P2 escape mutations, which from study of 043-C is not the case, the likelihood is that these particular KK10 escape mutants were transmitted, and did not arise independently ($P = 0.01$, Fisher's exact test).

We next evaluated the effect of the P2 anchor mutation in the KK10 epitope on the subsequent immune response in the children. In the three children infected with the non-binding KK10 variants, the dominant CTL specificity was still HLA-B27-restricted; however, it was directed against an epitope in p17 Gag and mapped to residues 19–27, sequence IRLRPGGKK (designated IK9, see Supplementary Information Fig. 1a). This response is typically undetectable or subdominant in adults with HLA-B27 (ref. 4) but evidently may become dominant when escape from KK10 occurs, such as in mother 048-M (see Supplementary Information Fig. 1b). All three children who lacked KK10 responses targeted this epitope. However, in child 043-C, who inherited B27 from the father, the KK10 response was clearly dominant (1,925 compared with 192 spot-forming cells (SFCs) per million peripheral blood mononuclear cells (PBMCs) specific for KK10 and IK9, respectively).

To determine the evolutionary origin of the mutations in the KK10 epitope, we performed phylogenetic analysis of the gag sequences in these subjects. Although the close genetic relationship of the viruses was evident within individual family groups, the intra-individual clustering confirmed divergent evolution of viruses after transmission (Fig. 3). Thus the sequence data presented exclude sample contamination and are best explained by transmission of the escape variants rather than multiple convergent evolution.

We next evaluated the stability of these CTL escape mutations. After 18 years of viral evolution in one family (10 years in 002-M and 8 years in 002-C) and 19 years in a second family (048-C, 049-C and 048-M) after transmission, the same rarely reported sequences are still shared between the mother and the child. As the mutations in the B27-KK10 viruses prevent binding, no CTL responses would be induced to maintain CTL-mediated immune selection pressure against this epitope. To test this further, viruses from the three children with CTL escape mutations were passaged 13–16 times in culture over a period of six months. No mutation back to the wild-type sequence was seen in viruses showing mutations at the P2 position, confirming that these mutated viruses are both replication competent and stable (data not shown)¹².

These data indicate that evolution and mother–child transmission of stable CTL escape mutations occurs for a dominant CTL response that is restricted by HLA-B27. To address whether selection occurs in the context of other epitopes and alleles, we examined

epitopes known to be targeted at the time of decreasing viral load in acute infection, including A3, B58 and B27 (refs 7, 13, 14). Viral sequences obtained from subjects with chronic infection were compared to the consensus sequence for each epitope⁸. Mutations in the dominant A3-restricted Gag epitope were observed in ≥90% of autologous clones in 17 out of 20 subjects who were A3-positive, but in only 7 out of 24 A3-negative subjects ($P = 0.0002$). Similarly, significant associations between expression of the relevant allele and mutation within the dominant epitope targeted were observed for the context of HLA-B58 ($P = 0.02$) and -B27 ($P = 0.0005$). In contrast, for a dominant A2-restricted Gag epitope that first appears long after the initial drop in viraemia¹³, there was no association between expression of A2 and mutation within the epitope in chronically infected patients^{14,15}.

Next, we investigated whether these epitope mutations are being sexually transmitted in adult infection. We examined adult subjects identified (since 1997) with acute HIV infection—defined as pre-seroconversion or within 180 days of infection—who did not express the relevant alleles, to be sure that epitope-specific CTL pressure would not be present. Viruses isolated from early infection encoding the relevant mutations were observed in 2 out of 15 B27-negative patients, 4 out of 10 A3-negative patients, and 4 out of 14 B58-negative patients with recent infection. Moreover, in one transmitting couple whose viral sequences are closely linked by phylogeny (subjects 6007 and 6008 in Fig. 3), both the B58-positive and the B58-negative subjects had viruses encoding mutations within the dominant B58-restricted epitope. The likelihood of a mutation in this epitope being present in both subjects by chance, using the frequencies described in the database⁸, would be 0.006. We thus conclude that viruses containing mutations in frequently targeted CTL epitopes are currently being sexually transmitted.

These data show that CTL selection pressure can influence virus

Table 1 Transmission and maintenance of CTL escape variants in children with HLA-B27

Subject	B27	KRWIILGLNK	Clones	IRLRPGGKK	Clones
043-M	–	-----M-----	4/4	-----S-----	4/4
043-C	+	-----M-----	3/3	-----S-----	3/3
048-M	+	-T---M-----	18/18	-----S-----	18/18
048-C	+	-T---M-----	11/11	-----S-----	6/11
				-----R-----	4/11
				-----RR-----	1/11
048-M	+	-T---M-----	18/18	-----S-----	18/18
049-C	+	-T---M-----	19/19	-----S-----	18/19
				-----K-----	1/19
002-M	+	-T---M-----	10/10	-----S-----	9/10
				-----G-----S-----	1/10
002-C	+	-T---M-----	8/8	-----S-----	7/8
				-----G-----S-----	1/8

Deduced, autologous amino-acid sequences for the B27-KK10 and IK9 epitopes in the B27-positive children, and mothers of these children. Sequences containing STOP codons were excluded (compare with Fig. 3).

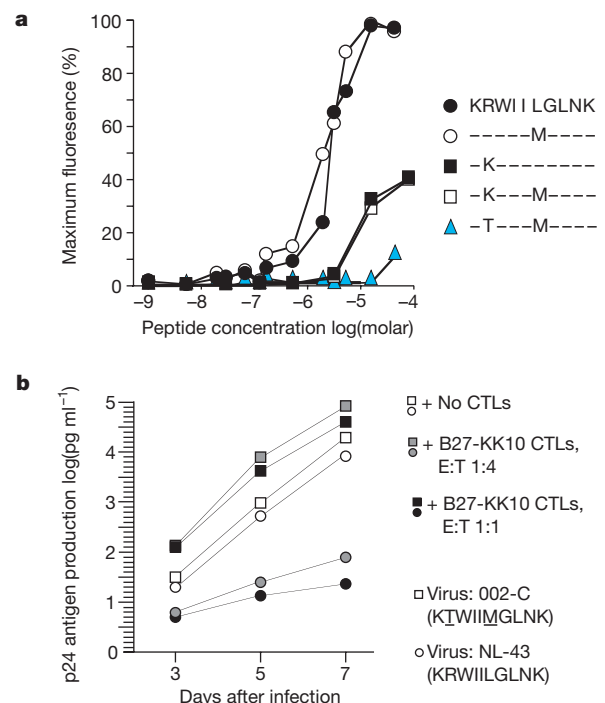


Figure 2 Binding and CTL recognition of the described KRWIILGLNK epitope variants. **a**, Binding of HLA-B27-restricted KRWIILGLNK (KK10) variants (KRWIILGLNK peptides, where X indicates position of the substituted amino acid) to T2 cells expressing B2705. **b**, Replication inhibition assays³⁰ using a CTL clone from 043-C specific for the B27-KK10 response. The data shown are representative of 19 similar assays using viruses cultured from the subjects described herein. The CD4⁺ T cells were HLA-matched only through HLA-B*2705 with the effector cells.

evolution and transmission, and suggest that viruses containing mutations in CTL epitopes can accumulate as the epidemic progresses, particularly as loss of immune control leads to increased viraemia and enhanced transmission^{16,17}. The data are also consistent with the hypothesis that the formation of escape mutants before transmission may contribute to the more rapid progression to

disease that is seen in perinatally infected children, in whom at least half of the HLA alleles are shared with the transmitter. The critical nature of the acute immune response in determining the outcome from AIDS virus infection^{3,18,19} suggests that the unavailability of epitopes targeted in acute infection will have a detrimental impact on long-term immune control of HIV. Experiments in mice

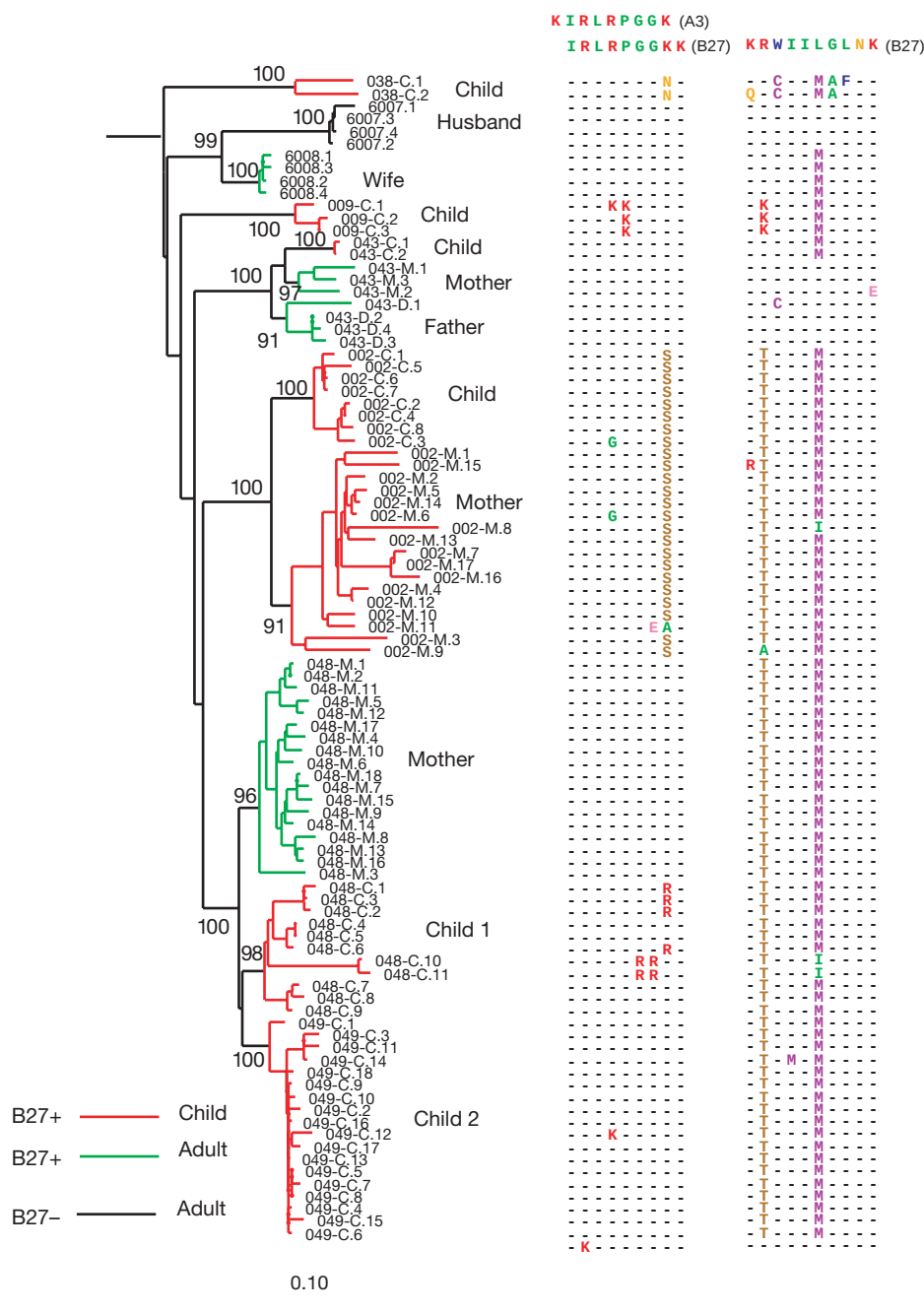


Figure 3 HLA-B27 epitope variants relative to the phylogenetic relationships of the HIV-1 sequences included in this study. The HIV-1 gag sequences from the study subjects were aligned with a subtype C sequence (92BR025, accession number U52953), which was used as an outgroup; the outgroup was trimmed from the tree to save space. The region encoding p17 and p24 was sequenced, and the alignment included 1,057 bases after gap stripping. Sequences from six children that are B27 positive are included, as well as their mothers' viral sequences where available, and in one case the father's. Both the bootstrap scores for within-patient clusters and the bootstrap scores for the associations between HIV-1 sequences derived from related patients are very high (always over 90%). Sequences from an HIV-positive married couple were also included in the tree. Epitope sequences are indicated at the top, and the most common form of the epitope in

the B clade is given as a reference; identity to the reference epitope is indicated by a dash. The substitutions in the KRWLILGLNK epitope are very rare in the background population. Two of the children that had potential escape mutations in position 8 of the B27 epitope IRLRPGGKK also expressed HLA-A3 (038-C and 002-C); escape from an A3-mediated response to the previously defined overlapping epitope KIRLRPGGK²⁶ might also be involved in these children. Amino acids are coloured according to a chemical classification scheme, indicating if a particular amino-acid substitution is conservative or not. Positive charge (K, R and H), red; negative charge (D and E), magenta; amide group (N and Q), orange; aliphatic (A, G, L, I, P and V), green; aromatic (F, W and Y), blue; sulphur containing (C and M), purple; hydroxylic (S and T), brown.

infected with lymphocytic choriomeningitis virus containing 1–3 mutated CTL epitopes demonstrated significantly impaired control of viraemia^{20,21}. Studies of sequence differences of A11-restricted Epstein–Barr virus-specific CTL epitope in populations where A11 is highly prevalent²² show lack of back reversion, as has been demonstrated for murine viruses that have escaped from neutralizing antibody responses²³, and here for CTL escape. The impact of this phenomenon in the evolving HIV epidemic is of critical relevance to vaccine design, as induction of CTLs that may be effective currently in controlling HIV-1 replication may be less relevant to the viruses circulating in the population in the future. Systematic evaluation will require a large investment in simultaneous characterization of immune responses and virus sequencing, particularly in rapidly expanding epidemics such as those currently occurring in Africa and Asia. □

Methods

Subjects studied

All six B27-expressing children were diagnosed as being infected with HIV-1 after parental diagnosis. None received antiretroviral therapy until over 1 year of age. (The viral loads measured at the time that highly active antiretroviral therapy was initiated, and other clinical parameters in adults and children, are shown in Supplementary Information Table 1.) Paediatric subjects were numbered consecutively, the suffix 'C' denoting the child studied, 'M' the mother of the child and 'D' the father.

HLA class I tissue typing and HLA-B*27 subtyping

HLA class I typing was performed by polymerase chain reaction single strand conformation polymorphism²⁴. The exception was 002-M whose cells were stained with the anti-B27-fluorescein isothiocyanate-conjugated antibody (Becton Dickinson) and analysed by flow cytometry.

Characterization of CTL responses

PBMCs in each of the subjects were screened for recognition in Elispot assays of epitopes within p17, p24, Nef, RT, gp41, gp120, Tat, Rev, Vif, Vpr and Vpu using overlapping peptides (OLPs) (refs 25, 26). In the children studied, where PBMC availability restricted screening, p17, p24 and Nef OLPs (002-C); p17, p24, Nef and RT OLPs (048-C and 049-C); and p17, p24, Nef, RT and gp41 OLPs (043-C) were used. In addition, B27-restricted epitopes defined in HIV²⁵ were also specifically tested.

Assays of CD8 cell function

We carried out Elispot assays as described²⁶. Background was ≤ 20 SFCs per million PBMCs (2 spots per well at 100,000 PBMCs per well) in all cases; responses of greater than 50 interferon (IFN)- γ SFCs per million PBMCs were significant. Responses >200 SFCs per million PBMCs were reconfirmed in intracellular IFN- γ staining assays. High frequency Elispot responses were quantified using lower input numbers of cells in repeat assays.

Intracellular cytokine staining for IFN- γ was performed as described elsewhere^{27,28} after stimulation of PBMCs with synthetic HIV peptide, anti-CD28 and anti-CD49d (Becton Dickinson) and addition 1 h later of Brefeldin A (Sigma). After a 6-h incubation at 37 °C, 5% CO₂, the cells were placed at 4 °C overnight. Cells were then washed and stained with surface antibodies anti-CD8 and anti-CD3 (Becton Dickinson), fixed and permeabilized (Caltag), and anti-IFN- γ monoclonal antibody was added (Becton Dickinson). Cells were then washed and analysed. Quadrant boundaries for IFN- γ staining were established by exclusion of greater than 99.97% of control CD8⁺ T cells. Epitope mapping, tetramer synthesis and analysis, and HLA restriction were performed as previously described^{26,28,29}.

Sequencing of viral DNA and sequence analyses

Genomic DNA was extracted from PBMC pellets using the Puregene DNA isolation kit (Gentra). HIV Gag sequences were amplified by nested PCR as described¹⁴. All PCR products were sequenced directly to determine the predominant viral quasiespecies present. In addition, cloned viral sequences were determined as previously described¹⁴, using the Topo2 cloning kit (Invitrogen) and the QiagenTurbo DNA purification kit. The phylogenetic tree in Fig. 3 was generated using the suite of programs in the PHYLIP package for creating neighbour-joining trees with the ML model and for generating bootstraps (<http://evolution.genetics.washington.edu/phyml.html>). The epitopes described in the text are B27-KK10 (KRWIIILGLNK, p24 GagHXB residues 131–140), B27-IK9 (IRLRPGGKK, p17 GagHXB residues 19–27), A3-RK9 (RLRPGGKKK, p17 GagHXB residues 20–28) and B58-TW10 (TSTLQEQIGW, p24 GagHXB residues 108–117)^{15,25}.

B*2705 binding assays

An HLA class I stabilization assay was used⁴, using peptide transporter (TAP)-deficient T2 cells transfected with B*2705. Cell-surface expression of B*2705 was measured with ME.1 (specific for HLA-B27, HLA-B7 and Bw22). We assessed maximum peptide binding using the peptide SRYWAIRTR (100 μ M). The relative binding of peptides is expressed as the per cent maximum change in fluorescence determined as follows: per cent maximum

change in fluorescence = $100 \times (\text{MFI for test peptide} - \text{MFI for diluent control}) / (\text{MFI for SRYWAIRTR} - \text{MFI for diluent control})$, where MFI is the mean fluorescence intensity.

HIV replication inhibition assays

CTL clones were generated by anti-CD3 stimulation of PBMCs at limiting dilution¹³. CD4⁺ T cells were expanded using PBMCs from an HIV-uninfected HLA-B*2705-positive donor to 100% purity over 6 d using an anti-CD8/CD3 bispecific antibody³⁰. Viruses cultured from the study subjects were used to infect 3×10^6 B27-expressing CD4⁺ T cells at a multiplicity of infection of 10:1 for 4 h. CTL clones were added to a total volume of 2 ml in a 24-well plate and placed in an incubator at 37 °C, 5% CO₂. Supernatant p24 antigen concentration was then determined.

Received 13 February; accepted 15 May 2001.

- Schmitz, J. E. *et al.* Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes. *Science* **283**, 857–860 (1999).
- Allen, T. M. *et al.* Tat-specific cytotoxic T lymphocytes select for HIV escape variants during resolution of primary viraemia. *Nature* **407**, 386–390 (2000).
- Rosenberg, E. S. *et al.* Immune control of HIV-1 after early treatment of acute infection. *Nature* **407**, 523–526 (2000).
- Goulder, P. J. *et al.* Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nature Med.* **3**, 212–217 (1997).
- Kelleher, A. D. *et al.* Clustered mutations in HIV-1 Gag are consistently required for escape from HLA-B27 restricted CTL responses. *J. Exp. Med.* **193**, 375–386 (2001).
- Phillips, R. E. *et al.* Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* **354**, 453–459 (1991).
- Wilson, J. D. *et al.* Direct visualization of HIV-1-specific cytotoxic T lymphocytes during primary infection. *Aids* **14**, 225–233 (2000).
- Human Retroviruses and AIDS 1999. *A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences* (ed Kuiken, C. L.) (Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico (<http://hiv-web.lanl.gov/HTML/99compendium.html>)), (1999).
- Kaslow, R. A. *et al.* Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nature Med.* **2**, 405–411 (1999).
- Barnhart, H. X. *et al.* Natural history of human immunodeficiency virus disease in perinatally infected children: an analysis from the Pediatric Spectrum of Disease Project. *Pediatrics* **97**, 710–716 (1996).
- Guo, H. C. *et al.* Comparison of the P2 specificity pocket in three human histocompatibility antigens: HLA-A*6801, HLA-A*0201, and HLA-B*2705. *Proc. Natl Acad. Sci. USA* **90**, 8053–8057 (1993).
- Nietfield, W. *et al.* Sequence constraints and recognition by CTL of an HLA-B27-restricted HIV-1 gag epitope. *J. Immunol.* **154**, 2189–2197 (1995).
- Altfield, M. *et al.* Cellular immune responses and viral diversity in individuals treated during acute and early HIV-1 infection. *J. Exp. Med.* **193**, 169–180 (2001).
- Goulder, P. J. *et al.* Substantial differences in specificity of HIV-specific cytotoxic T cells in acute and chronic HIV infection. *J. Exp. Med.* **193**, 181–194 (2001).
- Brander, C. *et al.* Lack of strong immune selection pressure by the immunodominant, HLA-A*0201-restricted cytotoxic T lymphocyte response in chronic human immunodeficiency virus-1 infection. *J. Clin. Invest.* **101**, 2559–2566 (1998).
- Connor, E. M. *et al.* Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. Pediatric AIDS Clinical Trials Group Protocol 076 Study Group. *N. Engl. J. Med.* **331**, 1173–1180 (1994).
- Quinn, T. C. *et al.* Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. *N. Engl. J. Med.* **342**, 921–929 (2000).
- Pantaleo, G. *et al.* The qualitative nature of the primary immune response to HIV infection is a prognosticator of disease progression independent of the initial level of plasma viremia. *Proc. Natl Acad. Sci. USA* **94**, 254–258 (1997).
- Barouch, D. H. *et al.* Control of viremia and prevention of clinical AIDS in rhesus monkeys by cytokine-augmented DNA vaccination. *Science* **290**, 486–492 (2000).
- Moskophidis, D. & Zinkernagel, R. M. Immunobiology of cytotoxic T-cell escape mutants of lymphocytic choriomeningitis virus. *J. Virol.* **69**, 2187–2193 (1995).
- Lewicki, H. A., Von Herrath, M. G., Evans, C. F., Whitton, J. L. & Oldstone, M. B. CTL escape viral variants. II. Biologic activity *in vivo*. *Virology* **211**, 443–450 (1995).
- de Campos-Lima, P. O. *et al.* HLA-A11 epitope loss isolates of Epstein–Barr virus from a highly A11+ population. *Science* **260**, 98–100 (1993).
- Ciurea, A., Hunziker, L., Klennerman, P., Hengartner, H. & Zinkernagel, R. Impairment of CD4(+) T cell responses during chronic virus infection prevents neutralizing antibody responses against virus escape mutants. *J. Exp. Med.* **193**, 297–306 (2001).
- Bunce, M. *et al.* Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). *Tissue Antigens* **46**, 355–367 (1995).
- Brander, C. & Goulder, P. J. R. in *HIV Molecular Immunology Database* (eds Korber, B. *et al.*) (Los Alamos National Laboratory, Los Alamos, New Mexico, 1999).
- Goulder, P. J. *et al.* Differential narrow focusing of immunodominant human immunodeficiency virus gag-specific cytotoxic T-lymphocyte responses in infected African and caucasoid adults and children. *J. Virol.* **74**, 5679–5690 (2000).
- Pitcher, C. J. *et al.* HIV-1-specific CD4+ T cells are detectable in most individuals with active HIV-1 infection, but decline with prolonged viral suppression. *Nature Med.* **5**, 518–525 (1999).
- Goulder, P. J. *et al.* Functionally inert HIV-specific cytotoxic T lymphocytes do not play a major role in chronically infected adults and children. *J. Exp. Med.* **192**, 1819–1832 (2000).
- Altfield, M. A. *et al.* Identification of dominant optimal HLA-B60- and HLA-B61-restricted cytotoxic T-lymphocyte (CTL) epitopes: rapid characterization of CTL responses by enzyme-linked immunospot assay. *J. Virol.* **74**, 8541–8549 (2000).
- Yang, O. O. *et al.* Suppression of human immunodeficiency virus type 1 replication by CD8+ cells: evidence for HLA class I-restricted triggering of cytolytic and noncytolytic mechanisms. *J. Virol.* **71**, 3120–3128 (1997).

Supplementary information is available at *Nature's* World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank those who have facilitated the collection of blood samples for this study, in particular N. Karthas, C. Kneut, S. Theodore and M. Phillips. We also thank J. Mear for help with binding assays, and the Women and Infant Transmission Study (WITS) for access to cryopreserved samples. The work was supported by the Medical Research Council (UK), the NIH, the Doris Duke Charitable Foundation, the Lloyd Foundation, the Elizabeth Glaser Pediatric AIDS Foundation, and a number of private donors. P.J.R.G. is an Elizabeth Glaser Scientist and B.D.W. is a Doris Duke Distinguished Clinical Science Professor.

Correspondence and requests for materials should be addressed to P.J.R.G. (e-mail: goulder@helix.mgh.harvard.edu).

An open form of syntaxin bypasses the requirement for UNC-13 in vesicle priming

Janet E. Richmond*, Robby M. Weimer* & Erik M. Jorgensen

Department of Biology, University of Utah, 257 South 1400 East, Salt Lake City, Utah 84112-0840, USA

* These authors contributed equally to this work

The priming step of synaptic vesicle exocytosis is thought to require the formation of the SNARE complex, which comprises the proteins synaptobrevin, SNAP-25 and syntaxin^{1–3}. In solution syntaxin adopts a default, closed configuration that is incompatible with formation of the SNARE complex⁴. Specifically, the amino terminus of syntaxin binds the SNARE motif and occludes interactions with the other SNARE proteins. The N terminus of syntaxin also binds the presynaptic protein UNC-13 (ref. 5). Studies in mouse, *Drosophila* and *Caenorhabditis elegans* suggest that UNC-13 functions at a post-docking step of exocytosis, most likely during synaptic vesicle priming^{6–8}. Therefore, UNC-13 binding to the N terminus of syntaxin may promote the open configuration of syntaxin⁹. To test this model, we engineered mutations into *C. elegans* syntaxin that cause the protein to adopt the open configuration constitutively⁴. Here we demonstrate that the open form of syntaxin can bypass the requirement for UNC-13 in synaptic vesicle priming. Thus, it is likely that UNC-13 primes synaptic vesicles for fusion by promoting the open configuration of syntaxin.

Using NMR spectroscopy, it has been shown that syntaxin adopts a closed configuration in solution⁴. However, mutations in two highly conserved amino acids (L165A, E166A) cause syntaxin to adopt a constitutively open configuration *in vitro*⁴. We made the corresponding mutations in *C. elegans* syntaxin (L166A, E167A; open syntaxin). Similar to vertebrate open syntaxin, the mutated *C. elegans* protein can bind synaptobrevin but not UNC-18 in pull-down assays (data not shown).

Expression of open syntaxin can fully rescue null mutations of syntaxin. *unc-64(js115)* is a null allele of the gene encoding *C. elegans* syntaxin¹⁰. Homozygotes of *unc-64(js115)* are completely paralysed and arrest development after hatching (Fig. 1a). This developmental defect was fully rescued by expression of wild-type syntaxin or the open form of syntaxin in null mutants (Fig. 1a). Expression of either form of syntaxin rescued the behavioural phenotypes associated with *unc-64(js115)* (Fig. 1b). Furthermore, expression of open syntaxin did not affect neuronal development (data not shown).

To test if open syntaxin could restore synaptic vesicle fusion we

quantified whole-cell voltage-clamp currents recorded from the neuromuscular junctions of dissected nematodes¹¹. The frequency of endogenous synaptic events in transgenic animals overexpressing wild-type syntaxin or open syntaxin were comparable ($P > 0.05$) to those in wild-type animals (Fig. 1c). In addition, stimulation of the presynaptic nerve cord in animals expressing wild-type or open syntaxin resulted in evoked release amplitudes that were not significantly different ($P > 0.05$) from those of wild-type animals (Fig. 1d). These data demonstrate that open syntaxin can substitute for wild-type syntaxin in basal synaptic transmission.

It was possible that these mutations in syntaxin created a gain-of-function protein that executes vesicle fusion independent of normal synaptic mechanisms. To demonstrate that open syntaxin functioned by means of the normal mechanisms of exocytosis, we confirmed that open syntaxin mediates exocytosis through the SNARE complex and requires calcium influx. We made a double mutant with open syntaxin and the synaptobrevin null allele *snb-1(js124)*¹². The larval lethality of *snb-1(js124)* was not rescued by open syntaxin ($n = 25$). To determine if open syntaxin could bypass the requirement of the N-type calcium channel UNC-2, we generated double mutants with *unc-2(e55)*¹³. Similarly, open syntaxin did not suppress the behavioural and neurotransmission defects associated with *unc-2* mutants (Fig. 2). These data demonstrate that open syntaxin functions through the normal synaptic machinery and validates it as a reagent to test the function of UNC-13.

Electrophysiological studies in mouse, *Drosophila* and *C. elegans* indicate that UNC-13 is probably required to prime synaptic vesicles for exocytosis^{6–8}. Because UNC-13 binds the N terminus of syntaxin⁵, it has been suggested that UNC-13 might promote the open state of syntaxin to enable vesicle priming⁹. If this model were correct then expression of open syntaxin should bypass the requirement of UNC-13 in synaptic vesicle exocytosis. To test this model we expressed the constitutively open form of syntaxin in *unc-13(s69)* animals. *unc-13(s69)* is a frameshift mutation before the highly conserved carboxy terminus of UNC-13 (ref. 14) and causes an almost complete paralysis in mutant nematodes (Fig. 3a).

Overexpression of wild-type syntaxin failed to rescue any of the *unc-13(s69)* behavioural phenotypes including locomotory movements, pharyngeal pumping and defecation cycles (Fig. 3a, b). However, overexpression of open syntaxin in *unc-13(s69)* animals caused a significant ($P < 0.0001$) improvement in all three behaviours (Fig. 3a, b). The suppression of the *unc-13* behavioural phenotypes suggested that open syntaxin was able to bypass the requirement for UNC-13 in synaptic transmission.

As almost no synaptic vesicle fusion is observed in *unc-13(s69)* animals, previous studies could not exclude UNC-13 from having a function in membrane fusion^{6–8}. If UNC-13 mediates fusion of the vesicular and plasma membranes, then open syntaxin should not be able to bypass this requirement for UNC-13 in fusion. To directly assay whether synaptic vesicles can fuse with the plasma membrane in the absence of UNC-13, we recorded spontaneous, miniature postsynaptic currents in the absence of calcium. In *unc-13(s69)* animals there was essentially no synaptic vesicle fusion in 0 mM calcium (Fig. 3c). Overexpression of wild-type syntaxin did not rescue vesicle fusion in *unc-13(s69)* animals. However, overexpression of open syntaxin completely restored spontaneous fusion in *unc-13(s69)* animals to wild-type levels. These data demonstrate that UNC-13 is not required at the fusion step during exocytosis.

UNC-13 contains several C2 domains that are calcium-binding motifs^{14–16}. The presence of these domains suggested that UNC-13 might be a calcium sensor for synaptic vesicle exocytosis. If UNC-13 were the sole calcium sensor, then in the absence of UNC-13 there should be no calcium-dependent release. Consistent with this hypothesis, *unc-13(s69)* mutants completely lack calcium-dependent evoked responses (Fig. 3e), and overexpression of wild-type syntaxin failed to rescue evoked release in the *unc-13(s69)* mutants. However, overexpression of open syntaxin completely restored

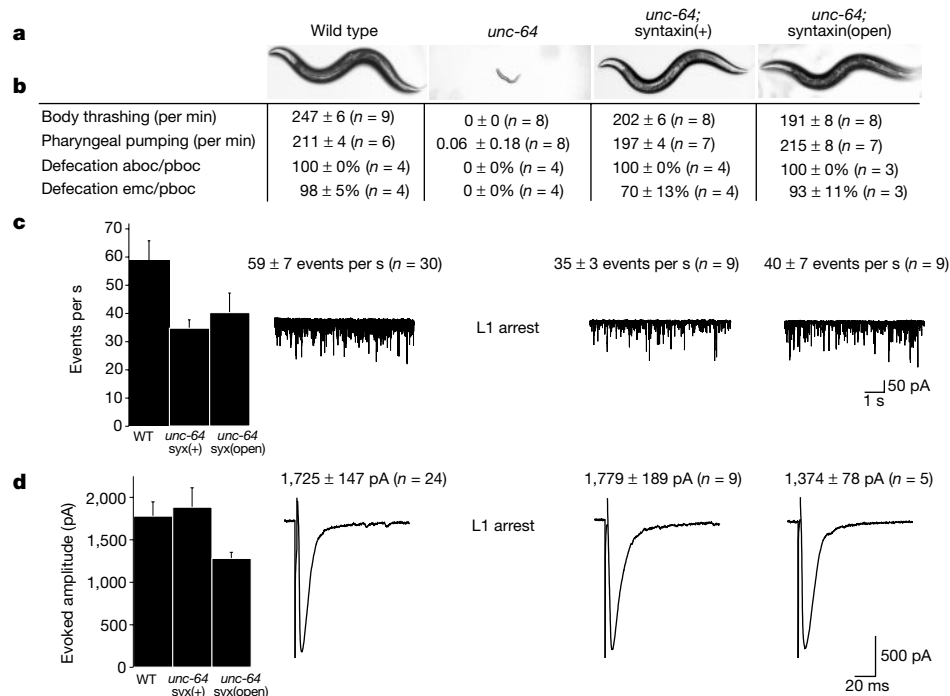


Figure 1 Expression of the constitutively open form of syntaxin rescued the syntaxin null *unc-64(j115)*. **a**, Transformation with either wild-type syntaxin (+) or open syntaxin (open) restored viability to *unc-64(j115)*. **b**, Behavioural phenotypes of *unc-64(j115)* animals overexpressing either wild-type syntaxin or open syntaxin are comparable to wild-type animals. Strains overexpressing either form of syntaxin are slightly sluggish on plates, as reflected in significantly lower thrashing rates ($P < 0.002$). This behavioural deficit is a consequence of syntaxin overexpression and not incomplete rescue, as animals that are wild type for *unc-64* overexpressing either open syntaxin or wild-type

syntaxin have the same phenotype (data not shown). aboc, anterior body contraction; pboc, posterior body contraction; emc, enteric muscle contraction. **c**, Endogenous release rates in 5 mM Ca^{2+} saline were restored to wild-type levels ($P > 0.05$) in both *unc-64*; syntaxin(+) and *unc-64*; syntaxin(open) animals. **d**, Depolarization of the ventral nerve cord elicited large evoked post-synaptic currents not significantly different from the wild type ($P > 0.05$) in both *unc-64*; syntaxin(+) and *unc-64*; syntaxin(open) animals. In all figures: left, recordings in 5 mM Ca^{2+} plotted as mean ± s.e.m.; right, sample traces.

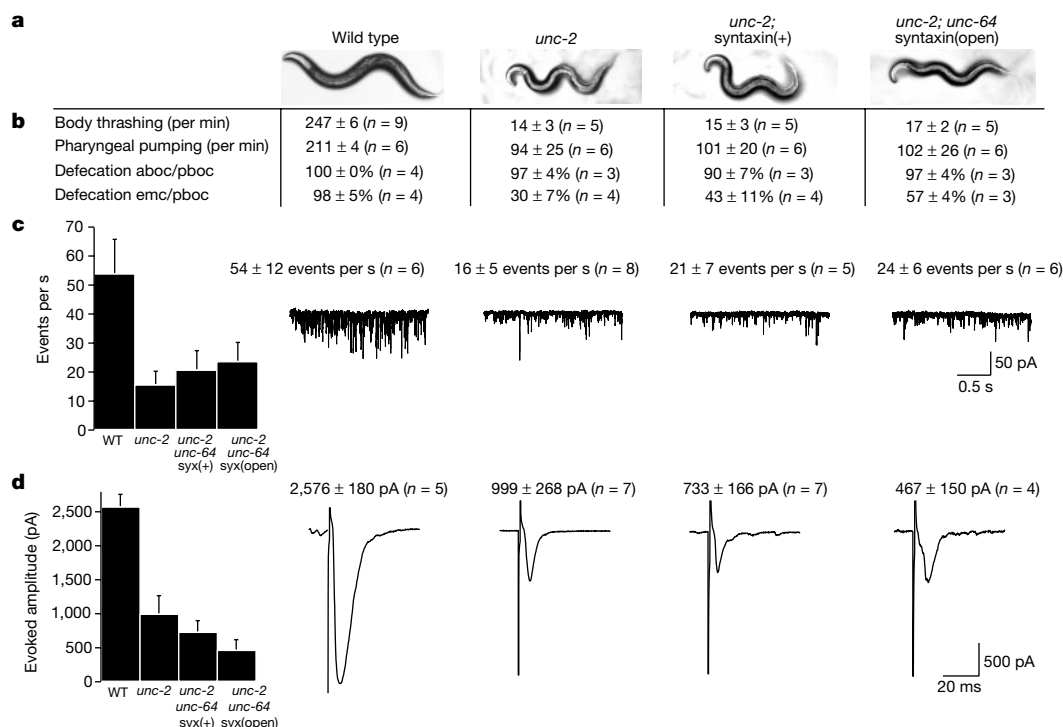


Figure 2 Open syntaxin failed to rescue mutants of the α subunit of the N-type voltage-gated calcium channel, UNC-2. **a**, The loss-of-function allele *unc-2(e55)* results in an uncoordinated locomotory phenotype. Expression of either syntaxin(+) or syntaxin(open) failed to restore locomotion to *unc-2(e55)* animals. **b**, Behavioural phenotypes associated with *unc-2* are not suppressed by the overexpression of syntaxin(+) or syntaxin(open). **c**, Loss of the UNC-2 calcium channel results in a significant

decrease in endogenous fusion events in 5 mM Ca^{2+} compared with the wild type ($P < 0.005$). Overexpression of syntaxin(+) or syntaxin(open) failed to restore wild-type, calcium-dependent endogenous currents in *unc-2* animals ($P < 0.05$). **d**, Loss of UNC-2 function results in a significant decrease in calcium-dependent evoked response compared with the wild type ($P < 0.002$). Overexpression of wild-type or open syntaxin failed to rescue the evoked response in *unc-2* animals ($P < 0.0001$).

evoked responses to wild-type levels (Fig. 3e). These normal responses to calcium in the absence of UNC-13 are consistent with the observation that overexpression of rat UNC-13 in chromaffin cells does not affect the calcium sensitivity of release¹⁷. Together, these data demonstrate that UNC-13 is not the calcium sensor that triggers fusion of synaptic vesicles. Previously we demonstrated that UNC-13 is essential for a post-docking step in synaptic vesicle exocytosis⁸. Similar results were obtained in null alleles of the *Drosophila* homologue, Dunc13 (ref. 7) and the mouse homologue, Munc13-1 (ref. 6). These studies concluded that UNC-13 probably has a role in priming of synaptic vesicles for fusion but they could not exclude a role for UNC-13 in either calcium sensing or in the fusion event itself. Our data demonstrate that UNC-13 is not required for calcium sensing or for fusion and strongly suggest a highly specific role for UNC-13 in the priming step of synaptic vesicle exocytosis.

Vesicles become fusion competent at the priming step of exocytosis. At the molecular level, priming is thought to be mediated by the formation of the SNARE complex^{1–3}. Overexpression of Munc13-1 in bovine chromaffin cells accelerates the forward rate constant for the priming of morphologically docked, large dense-core vesicles without affecting the rate of fusion or the calcium sensitivity of release¹⁷. This stage of dense-core vesicle exocytosis coincides with the association of the SNARE proteins¹⁸. Although Munc13-1 levels are normally very low in chromaffin cells, these observations suggest that UNC-13 can function to promote dense-core vesicle priming, possibly by promoting formation of the SNARE complex. Our data confirm and extend these studies by

demonstrating that UNC-13 promotes the priming of synaptic vesicles by acting through syntaxin. Specifically, the role of UNC-13 may be to bind the autoinhibitory domain of syntaxin to promote or maintain the open state and thus facilitate formation of the SNARE complex.

The rescue of the *Unc-13* phenotype from no evoked responses to wild-type levels of evoked responses by open syntaxin is a dramatic result; however, these animals are not completely wild type. First, body thrashing and locomotory activity of the mutant is greatly reduced compared with the wild type (Fig. 3b). Second, measures of endogenous release of synaptic vesicles in the presence of calcium is also greatly reduced compared with the wild type (Fig. 3d). There are several possible explanations for these results. One potential explanation is that the L166A and E167A mutations do not completely mimic the conformation of syntaxin when it is bound to UNC-13. Alternatively, UNC-13 may have an additional role in vesicle exocytosis, possibly to tether synaptic vesicles near calcium channels. Nevertheless, these data suggest that UNC-13 stimulates priming by opening syntaxin either through the direct interaction previously demonstrated or by acting on another protein, such as UNC-18. □

Methods

Genetics

We grew strains on nematode growth media (NGM)¹⁹. The wild-type strain is N2 Bristol. The *unc-13(s69)* allele is a putative null and contains a 5-base-pair deletion, which creates a frameshift at residue 1,029 (ref. 14). The syntaxin *unc-64(js115)* allele contains a nonsense mutation at codon 71 (ref. 10). In all cases the syntaxin(+) strain is EG1986 *unc-64*

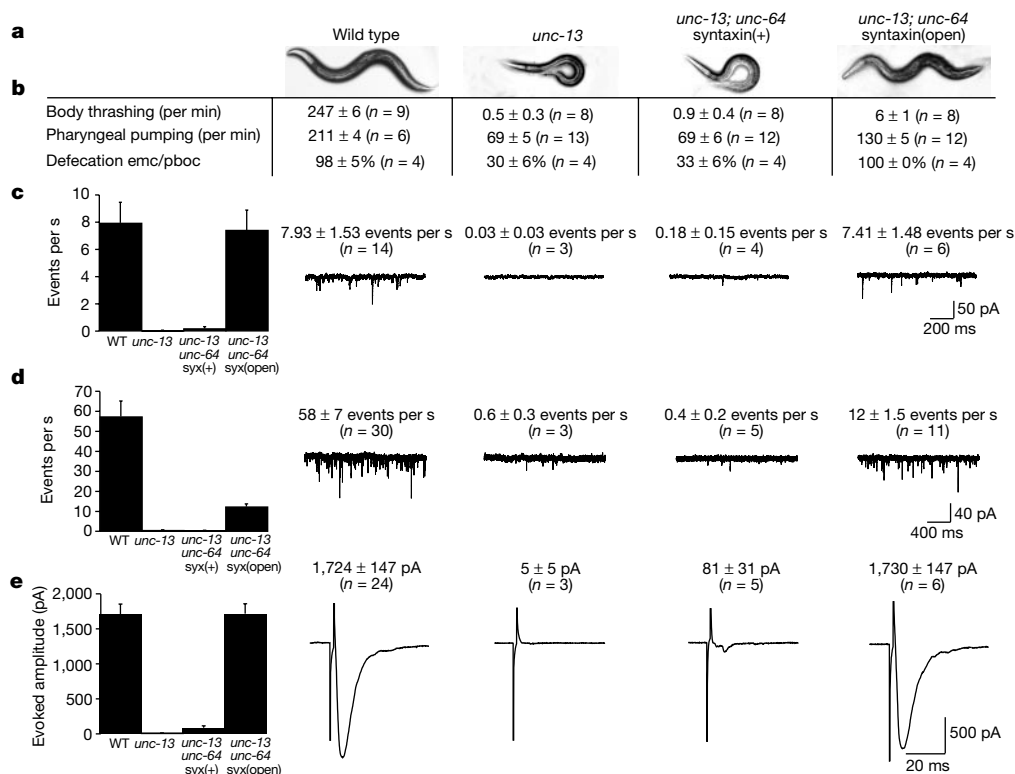


Figure 3 The open form of syntaxin rescues the *unc-13(s69)* mutant. **a**, Overexpression of open syntaxin, but not wild-type syntaxin, partially suppresses *unc-13* paralysis.

b, Expression of the open form of syntaxin significantly suppresses the body thrashing ($P < 0.0005$), pharyngeal pumping ($P < 0.0001$) and the defecation defects ($P < 0.0001$) associated with *unc-13(s69)*. **c**, Synaptic vesicle fusion can occur in the absence of UNC-13. The average spontaneous event frequency in 0 mM Ca^{2+} of *unc-13(s69); syntaxin(open)* animals was not significantly different ($P > 0.8$) from that of the wild type. **d**, Expression of open syntaxin partially restores endogenous vesicle fusion events in *unc-13(s69)* animals. The endogenous event frequency recorded in 5 mM Ca^{2+}

was significantly greater in *unc-13(s69); syntaxin(open)* than in both *unc-13(s69)* ($P < 0.005$) or *unc-13(s69); syntaxin* animals ($P < 0.0005$), but significantly lower than wild-type levels ($P < 0.0001$). **e**, Open syntaxin rescues calcium-dependent evoked responses in 5 mM Ca^{2+} in the *unc-13(s69)* mutant. Ventral cord stimulation elicited robust currents at the neuromuscular junctions of *unc-13(s69); syntaxin(open)* animals similar to those in the wild type. The average amplitude of the evoked responses in *unc-13(s69); syntaxin(open)* animals was not statistically different ($P > 0.05$) from that of wild-type animals.

(*js115*); *oxIs33[unc-64(+); Punc-122::GFP]* I. The syntaxin(open) strain is EG1985 *unc-64(js115) oxIs34[unc-64(L166A/E167A); Pmyo-2::GFP]* III. We generated the syntaxin(+) strain by microinjecting pTX21, a subclone of the *unc-64* locus¹⁰, into the germline of NM979 hermaphrodites (genotype: *unc-64(js115)/bli-5(e518)*) along with plasmid pPD97/98(*Punc-122::GFP*)²⁰ as a dominant co-injection marker. Green fluorescent protein (GFP+) transgenic progeny were assayed for rescue of *unc-64(js115)* segregants. Three out of three lines carrying pTX21 were fully rescued. Stably transmitted rescuing arrays were integrated into the genome by X-ray irradiation. The syntaxin(open) strain was generated in a similar fashion by microinjecting pJR04, a variant of pTX21 containing the corresponding L166A/E167A mutations in the *unc-64* coding region, along with plasmid pPD118.33 (*Pmyo-2::GFP*) (1997 Fire vector kit) as the dominant co-injection marker. Two out of two lines carrying pJR04 were fully rescued. The *unc-13(s69); syntaxin(+)* strain is EG1983 *unc-13(s69) oxIs33; unc-64(js115)*. The *unc-13(s69); syntaxin(open)* strain is EG1984 *unc-13(s69); unc-64(js115) oxIs34*. The *snb-1* null allele is *js124* (ref. 12) and the *unc-2* loss-of-function allele is *e55* (ref. 19).

Molecular biology

Plasmid pJR04: *unc-64(L166A/E167A)* was generated from pTX21 using the QuikChange Site-Directed Mutagenesis Kit (Stratagene) and oligonucleotides oJR005 and oJR006. The sequence of pJR04 was confirmed using an Applied Biosystems automated DNA sequencer at the Sequencing Core Facility (University of Utah). Sequence of oligonucleotides: oJR005, ggagatgagatgcggcgaaatgattgagagcgg; oJR006, ccgctctcaatcattcggccatctctcatctcc.

Behavioural assays

Body thrashing, pharyngeal pumping in the presence of food and the defecation cycle were assayed as described^{21,22}. Overexpression of wild-type or open syntaxin in *unc-64(js115)* resulted in sluggish animals. This subtle phenotype was also observed in *unc-64(+)* strains overexpressing either wild-type or open syntaxin. Therefore, sluggish movement is due to syntaxin overexpression rather than incomplete rescue. Furthermore, several transgenic arrays and many integrants showed similar behaviours, thus the behavioural defects are not due to the specific composition of the array or the site of integration (data not shown).

Electrophysiology and worm dissection

We performed electrophysiological methods as described^{8,11}. Briefly, animals were immobilized with a cyanoacrylic glue and a lateral incision was made to expose the ventral medial body wall muscles. Electrophysiological recordings from muscles were made in the whole-cell voltage-clamp configuration (holding potential -60 mV) at room temperature (21 °C) using an EPC-9 patch-clamp amplifier (HEKA) and digitized at 2.9 kHz through an ITC-16 interface (Instrutech). Data were acquired by Pulse software (HEKA) run on a Power Mac 6500/225. The bath solution contained 150 mM NaCl, 5 mM KCl, 5 mM CaCl₂, 1 mM MgCl₂, 10 mM glucose and 15 mM HEPES, pH 7.35, sucrose to ~330 mosmol. The pipette solution contained: 120 mM KCl, 20 mM KOH, 4 mM MgCl₂, 5 mM N-tris (hydroxymethyl)methyl-2- aminoethane-sulphonic acid, 0.25 mM CaCl₂, 4 mM NaATP, 36 mM sucrose, 5 mM EGTA, pH 7.2, sucrose to 315 mosmol. Subsequent analysis and graphing were performed using Pulsefit (HEKA), Mini Analysis (Jaen Software) and Igor Pro (Wavemetrics). All statistically derived values are given as mean $\pm$ s.e.m.

Received 21 March; accepted 4 June 2001.

- Sollner, T. *et al.* SNAP receptors implicated in vesicle targeting and fusion. *Nature* **362**, 318–324 (1993).
- Broadie, K. *et al.* Syntaxin and synaptobrevin function downstream of vesicle docking in *Drosophila*. *Neuron* **15**, 663–673 (1995).
- Hanson, P. I., Roth, R., Morisaki, H., Jahn, R. & Heuser, J. E. Structure and conformational changes in NSF and its membrane receptor complexes visualized by quick-freeze/deep-etch electron microscopy. *Cell* **90**, 523–535 (1997).
- Dulubova, I. *et al.* A conformational switch in syntaxin during exocytosis: role of munc18. *EMBO J.* **18**, 4372–4382 (1999).
- Betz, A., Okamoto, M., Benseler, F. & Brose, N. Direct interaction of the rat *unc-13* homologue Munc13-1 with the N terminus of syntaxin. *J. Biol. Chem.* **272**, 2520–2526 (1997).
- Augustin, I., Rosenmund, C., Sudhof, T. C. & Brose, N. Munc13-1 is essential for fusion competence of glutamatergic synaptic vesicles. *Nature* **400**, 457–461 (1999).
- Aravamudan, B., Fergestad, T., Davis, W. S., Rodesch, C. K. & Broadie, K. *Drosophila* Unc-13 is essential for synaptic transmission. *Nature Neurosci.* **2**, 965–971 (1999).
- Richmond, J. E., Davis, W. S. & Jorgensen, E. M. UNC-13 is required for synaptic vesicle fusion in *C. elegans*. *Nature Neurosci.* **2**, 959–964 (1999).
- Brose, N., Rosenmund, C. & Rettig, J. Regulation of transmitter release by Unc-13 and its homologues. *Curr. Opin. Neurobiol.* **10**, 303–311 (2000).
- Saifee, O., Wei, L. & Nonet, M. L. The *Caenorhabditis elegans* *unc-64* locus encodes a syntaxin that interacts genetically with synaptobrevin. *Mol. Biol. Cell* **9**, 1235–1252 (1998).
- Richmond, J. E. & Jorgensen, E. M. One GABA and two acetylcholine receptors function at the *C. elegans* neuromuscular junction. *Nature Neurosci.* **2**, 791–797 (1999).
- Nonet, M. L., Saifee, O., Zhao, H., Rand, J. B. & Wei, L. Synaptic transmission deficits in *Caenorhabditis elegans* synaptobrevin mutants. *J. Neurosci.* **18**, 70–80 (1998).
- Schafer, W. R. & Kenyon, C. J. A calcium-channel homologue required for adaptation to dopamine and serotonin in *Caenorhabditis elegans*. *Nature* **375**, 73–78 (1995).
- Kohn, R. E. *et al.* Expression of multiple UNC-13 proteins in the *C. elegans* nervous system. *Mol. Biol. Cell* **11**, 3441–3452 (2000).
- Maruyama, I. N. & Brenner, S. A phorbol ester/diacylglycerol-binding protein encoded by the *unc-13* gene of *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **88**, 5729–5733 (1991).
- Brose, N., Hofmann, K., Hata, Y. & Sudhof, T. C. Mammalian homologues of *Caenorhabditis elegans* *unc-13* gene define novel family of C2-domain proteins. *J. Biol. Chem.* **270**, 25273–25280 (1995).

- Ashery, U. *et al.* Munc13-1 acts as a priming factor for large dense-core vesicles in bovine chromaffin cells. *EMBO J.* **19**, 3586–3596 (2000).
- Xu, T. *et al.* Inhibition of SNARE complex assembly differentially affects kinetic components of exocytosis. *Cell* **99**, 713–722 (1999).
- Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
- Miyabayashi, T., Palfreyman, M. T., Sluder, A. E., Slack, F. & Sengupta, P. Expression and function of members of a divergent nuclear receptor family in *Caenorhabditis elegans*. *Dev. Biol.* **215**, 314–331 (1999).
- Miller, K. G. *et al.* A genetic selection for *Caenorhabditis elegans* synaptic transmission mutants. *Proc. Natl Acad. Sci. USA* **93**, 12593–12598 (1996).
- Thomas, J. H. Genetic analysis of defecation in *Caenorhabditis elegans*. *Genetics* **124**, 855–872 (1990).

Acknowledgements

We thank M. Nonet for providing *unc-64(js115)*, RAB-3 antibodies and the plasmid pTX21; R. Hosono for providing UNC-18 antibodies; A. Rose for providing the *unc-13(s69)* allele; and P. Sengupta for the *Punc-122::GFP* plasmid. We also thank K. Broadie for critical review of this manuscript. This work was supported by NIH grants to J.E.R. and E.M.J.

Correspondence and requests for materials should be addressed to E.M.J. (e-mail: jorgensen@biology.utah.edu)

Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas

Laura Pasqualucci*, Peter Neumeister*, Tina Goossens†, Gouri Nanjangud‡, R. S. K. Chaganti‡, Ralf Küppers†* & Riccardo Dalla-Favera*

* Institute for Cancer Genetics and the Department of Pathology, Columbia University, New York, New York 10032, USA

† Institute for Genetics, University of Cologne, 50931 Cologne, Germany

‡ Laboratory of Cancer Genetics and the Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

Genomic instability promotes tumorigenesis and can occur through various mechanisms, including defective segregation of chromosomes or inactivation of DNA mismatch repair¹. Although B-cell lymphomas are associated with chromosomal translocations that deregulate oncogene expression², a mechanism for genome-wide instability during lymphomagenesis has not been described. During B-cell development, the immunoglobulin variable (V) region genes are subject to somatic hypermutation in germinal-centre B cells³. Here we report that an aberrant hypermutation activity targets multiple loci, including the proto-oncogenes *PIM1*, *MYC*, *RhoH/TTF* (*ARHH*) and *PAX5*, in more than 50% of diffuse large-cell lymphomas (DLCLs), which are tumours derived from germinal centres. Mutations are distributed in the 5' untranslated or coding sequences, are independent of chromosomal translocations, and share features typical of V-region-associated somatic hypermutation. In contrast to mutations in V regions, however, these mutations are not detectable in normal germinal-centre B cells or in other germinal-centre-derived lymphomas, suggesting a DLCL-associated malfunction of somatic hypermutation. Intriguingly, the four hypermutable genes are susceptible to chromosomal translocations in the same region, consistent with a role for hypermutation in generating translocations by DNA double-strand breaks^{4–6}. By mutating multiple genes, and possibly by favouring chromosomal translocations, aberrant hypermutation may represent the major contributor to lymphomagenesis.

Somatic hypermutation occurs in germinal-centre B cells and is found in all germinal-centre-derived B-cell tumours³. This process

Table 1 Features of mutations in *PIM1*, *PAX5*, *RhoH/TF*, *MYC*, *VH* and *BCL6* in DLCL

Locus	Mutation frequency, per 100 bp (range)*	Single base-pair substitutions	Deletions and insertions	Transitions over transversions	RGYW bias, per 100bp†
<i>PIM1</i>	0.2 (0.04–0.7)	57	5	4.0	0.8 ($P < 0.001$)
<i>PAX5</i>	0.15 (0.06–0.3)	41	4	1.2	0.8 ($P < 0.01$)
<i>RhoH/TF</i>	0.17 (0.05–0.4)	44	3	1.3	0.36 ($P < 0.05$)
<i>MYC</i>	0.12 (0.03–0.6)	59	1	1.2	0.6 ($P < 0.001$)
	0.19 (0.06–0.4)				
<i>VH</i>	12.7 (1–23.2)	334	0	1.3	39 ($P < 0.001$)
<i>BCL6</i>	0.69 (0.06–2.9)	211	5	1.0	2.6 ($P < 0.01$)

The presence of strand polarity could not be assessed owing to the low number of mutational events.

* Mutation frequencies were calculated in the region containing >90% of the mutations on mutated cases only (Fig. 1, greyed area). For the *MYC* gene, values for both the exon 1–intron 1 and the exon 2 clusters are given.

† Frequency of mutated G bases occurring in the context of an RGYW motif. The statistical significance compared to the expected frequency of mutations, as calculated by the χ^2 test, is given in parentheses.

was thought to selectively target the rearranged *V* genes to generate antibody diversity. However, it has been shown that the 5' sequences of *BCL6* and *Fas/CD95* (*TNFRSF6*) are also mutated in normal germinal-centre B lymphocytes^{7–9}, suggesting that somatic hypermutation may target more genes than originally suspected. This process requires transcription and typically affects genomic sequences up to ~2 kb (kilobases) downstream from the site of transcription initiation¹⁰. Thus, to search for novel targets, we selected 18 genes, including *BCL6* as a control (see Supplementary Information Table 1 for a complete list), on the basis of their expression in germinal-centre B cells, then directly sequenced the products of polymerase chain reaction (PCR) generated from their 5' regions (~0.5–3 kb downstream to the promoter). Mutation screening was performed on DLCL samples (39 in total), as this tumour derives from malignant transformation of a cell that transited the germinal centre and therefore contains hypermutated immunoglobulin *V* genes.

Whereas germline sequences were found at 13 of the 18 loci examined in most DLCL cases, a significant fraction of these tumours displayed sequence variants in *PIM1*, *MYC*, *RhoH/TF* and *PAX5* (43, 32, 46 and 57%, respectively). As expected, the 5' non-coding region of *BCL6* was mutated in 71% of cases^{7,11}. Overall, >50% of DLCLs had mutations in at least two of these four genes (specifically, 4 in 11% of the cases, 3 in 21%, 2 in 25%, 1 in 18% and 0 in 25%), with a trend to display more mutations per gene in samples harbouring more mutated genes (Supplementary Information Table 2). In a subset of cases, sequencing of cloned PCR products showed that one or two predominant alleles recapitulate the mutations observed by direct sequencing, confirming their presence in the tumour clone and revealing their frequently bi-allelic distribution (Supplementary Information Fig. 1). Mutations were somatic in origin, as they were not present in normal DNA from the same patients ($n = 11$ cases). Cytogenetic analysis, including fluorescent *in situ* hybridization with probes for the immunoglobulin heavy chain locus, excluded the possibility that hypermutation was due to chromosomal translocation and linkage to immunoglobulin genes, as is the case for translocated *MYC* alleles in Burkitt lymphoma¹².

We found 62 mutations unique to individual tumour DNAs in 12 of 28 DLCL cases (43%) in *PIM1*, a proto-oncogene identified as a preferential proviral integration site in murine T-cell lymphomas¹³

and occasionally involved in DLCL-associated chromosomal translocations¹⁴. These mutations were distributed in a region spanning up to 2 kb towards the 3' end from the transcription initiation site, and were clustered in a stretch of ~1.2 kb (Fig. 1a, greyed area) that was found mutated in 11 of 12 positive cases (92%) and contained >90% of the mutations (average frequency 0.2 per 100 base pairs (bp) in the mutated cases; Table 1). Notably, seven missense mutations introduced into the coding exons of four cases (Ly8, 470, 1213 and 1576; see Supplementary Information) predict a change in the structure and, in some cases, the function of the *PIM1* protein, a serine/threonine protein kinase¹⁵ involved in cell proliferation and survival¹⁶.

A second gene found hypermutated was *MYC*, which encodes a transcription factor involved in the control of cell growth, proliferation, differentiation and apoptosis, and is widely implicated in tumorigenesis by a variety of mechanisms, including chromosomal translocations¹⁷. Tumour-associated mutations of the *MYC* gene have been observed in endemic Burkitt lymphoma carrying t(8;14) translocations, presumably owing to somatic hypermutation driven by immunoglobulin sequences juxtaposed to the *MYC* locus^{12,18}. In DLCL, we identified two clusters of mutations in 12 (32%) of 37 samples (Fig. 1b and Supplementary Information). With the exception of three cases, mutations were alternatively distributed in the region downstream to the major P1/P2 (exon 1) or the region downstream to the minor P3 promoter (exon 2)², consistent with the dependence of these mutations on transcription from one of these two promoters. The mutations observed within the exon 1–intron 1 cluster were analogous in frequency and distribution to those described in Burkitt lymphoma, which in some cases have been shown to alter *MYC* transcription by releasing a block on transcriptional elongation¹⁹. In three samples, eight missense mutations were found within the sequences of exon 2 that encode the transactivation domain (Supplementary Information Table 4b). Mutations in this region can deregulate *MYC* function by interfering with its phosphorylation, protein stability or repression of trans-activation activity by the RB-related protein p107^{18,20}. In three cases amino-acid replacements at the same residue had been previously reported in Burkitt lymphoma. These findings suggest that at least some of the translocation-independent mutations in DLCL have been selected for their ability to alter *MYC* expression or function.

Thirteen (46%) of 28 DLCLs exhibited mutations scattered

Table 2 Single-cell analysis of *RhoH/TF*, *Pax5*, *BCL6* and *VH* in normal B lymphocytes

Cell phenotype	Donor	Number of mutated cells*				Number of mutations (%)†			
		<i>RhoH/TF</i>	<i>PAX5</i>	<i>BCL6</i>	<i>VH</i>	<i>RhoH/TF</i>	<i>PAX5</i>	<i>BCL6</i>	<i>VH</i>
IgD ⁺ CD27 ⁺	I	2/20	1/11	1/21	1/14	2 (0.008)	2 (0.034)	1 (0.005)	3 (0.08)
	II	0/10	0/7	ND	0/6	0	0	ND	0
CD38 ⁺ CD77 ⁺	I	2/21	1/9	5/15	12/12	2 (0.008)	1 (0.020)	14 (0.10)	95 (4.2)
	II	0/13	2/11	ND	8/9	0	2 (0.034)	ND	100 (5)

ND, not determined.

* Number of cells harbouring sequence variants over total number of cells sequenced.

† Total number of mutations found. Based on the presence, in both the *RhoH/TF* and the *BCL6* gene from donor I, of two polymorphisms located on distinct alleles, it was calculated that on average 1.3 alleles per cell were amplified, corresponding to 1,137 bp per cell for the *RhoH/TF* gene, 533 bp per cell for the *PAX5* gene, and 930 bp per cell for *BCL6*.

throughout the first 1.6kb of the *RhoH/TTF* gene (Fig. 1c and Supplementary Information), which encodes a small GTP-binding protein belonging to the *RAS* superfamily²¹. This gene is involved in rare tumour-associated translocations that juxtapose its coding domain to the immunoglobulin locus, suggesting that *RhoH/TTF* is a proto-oncogene²¹. The sequence variants (average mutation

frequency 0.17 per 100 bp; Table 1) reside within non-coding sequences, suggesting a potential effect on regulation of *RhoH/TTF* gene expression.

The fourth gene found mutated in DLCL, *PAX5*, encodes a B-cell-specific transcription factor essential for B-lineage commitment and differentiation and involved in translocations in ~50% of

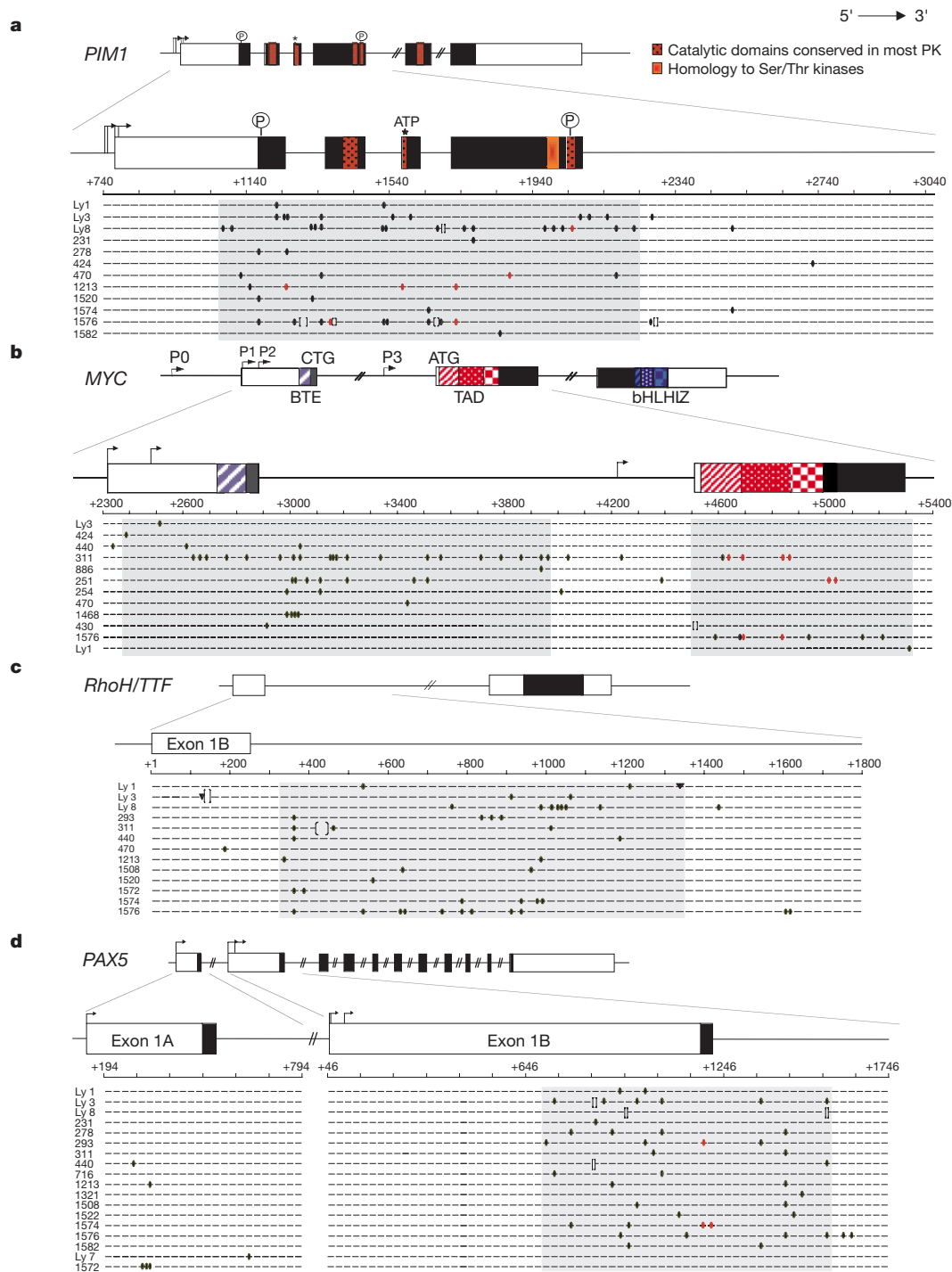


Figure 1 Mutational analysis of *PIM1* (a), *MYC* (b), *RhoH/TTF* (c) and *PAX5* (d). Genomic loci are shown with untranslated and translated sequences (open and filled boxes); hatched boxes are relevant protein functional domains. Arrows indicate transcriptional start sites. The regions amplified for analysis are expanded for each gene and aligned to sequences of mutated DLCL cases shown below (1 line = 2 alleles) where each small segment represents a 25-bp interval (numbered according to the GenBank accession numbers). Ovals, 1-bp substitutions (red for missense mutations); brackets, deletions;

triangles, insertions. Greyed areas represent the regions found mutated in >90% of the altered cases and containing >90% of the mutations. **a**, *PIM1*. ATP, site of ATP-binding; P, putative autophosphorylation sites. PK, protein kinases. **b**, *MYC*. The two alternative translational start sites are indicated by CTG and ATG. BTE, segment controlling block of transcriptional elongation; TAD, transactivation domain; bHLHIZ, basic helix-loop-helix leucine zipper domain.

lymphoplasmacytoid lymphoma²². Somatic mutations were identified downstream of both transcription initiation sites, predominantly around exon 1B (16 of 28 cases, 57%), at a frequency of 0.15 per 100 bp (Fig. 1d and Table 1). Although most of the changes involved non-coding sequences, three missense mutations were detected within the short coding portion of exon 1B.

To investigate whether mutations in these genes can be acquired by normal B cells during transit through the germinal centre, we analysed *PIM1*, *RhoH/TTF* and *PAX5*, along with *BCL6* and the immunoglobulin heavy chain V region (*V_H*) as controls, in normal B-cell subpopulations. Previous studies have already established the absence of *MYC* mutations in human post-germinal-centre memory B cells^{8,23}. First, naive B cells (IgD⁺CD27⁺) and germinal-centre centroblasts (CD38⁺CD77⁺) were individually sorted from the tonsils of two donors^{4,7} and subjected to primer-extension preamplification followed by single-cell PCR and sequencing⁹ at the *RhoH/TTF* and *PAX5* loci. As expected, gene analysis of *V_H* and *BCL6* showed somatically mutated sequences in germinal-centre, but not naive, B cells, validating the cell-purification procedure (Table 2). In contrast, only rare changes were found in *RhoH/TTF* and *PAX5*, with no significant differences between naive and germinal-centre cells (Table 2), suggesting that they may represent *Taq* DNA polymerase misincorporations during the preamplification step. To corroborate these results, we analysed *PIM1*, *RhoH/TTF*, *PAX5*, *BCL6* and *V_H* in purified naive and germinal-

centre B-cell populations by sequencing cloned PCR products obtained using the low-error Pfu polymerase²⁴. The results confirmed the lack of mutations in the *PIM1*, *RhoH/TTF* and *PAX5* sequences from germinal-centre lymphocytes, as well as naive B cells and control fibroblasts, in the presence of the expected levels of mutations in *BCL6* and *V_H* (Fig. 2a). Together with previous studies on *MYC*^{8,23}, our findings indicate that hypermutation of these four genes does not occur during normal B-cell development and thus represents a tumour-associated event.

To determine whether hypermutation of these four genes is restricted to DLCL or occurs in other B-cell malignancies, we screened a panel of B-cell non-Hodgkin's lymphoma (NHL), including pre-germinal-centre tumours lacking *V_H* and *BCL6* mutations, germinal-centre-derived lymphomas, and chronic lymphocytic leukemia (CLL) (see Fig. 2b legend). In these tumours, mutations were completely absent (*PIM1* and *MYC*) or rare (*RhoH/TTF* and *PAX5* in Burkitt lymphoma and follicular lymphoma), with the expected exception of *MYC* in endemic Burkitt lymphoma cases carrying the (8;14) translocation (Fig. 2b and Supplementary Information Table 5). The genes *Fas*, *L-Plastin* (*LCPI*) and *ICAM1*, analysed as controls since they are only rarely mutated in DLCL, were also found in germline configuration in follicular lymphoma, Burkitt lymphoma and multiple myeloma (data not shown). Overall, these results indicate that, in contrast to *V_H* and *BCL6*, hypermutation of *PIM1*, *MYC*, *RhoH/TTF* and *PAX5* is not common to all germinal-centre-derived tumours, but is instead largely restricted to DLCL. Alternatively, it is possible that different mutated genes are selected in different lymphoma types.

To investigate the mechanism underlying the DLCL-associated hypermutation, we examined whether specific features of the somatic hypermutation process were also detectable in the mutated *PIM1*, *MYC*, *RhoH/TTF* and *PAX5* genes. The average mutation frequency in the cluster regions was 3–6-fold and 50–100-fold lower than that observed in *BCL6* and *V_H*, respectively (Table 1); these frequencies probably result from the intrinsic mutation rate as well as from biological selection of mutants. The distribution of the mutations and its association with transcription initiation was remarkably similar to that reported for *V_H* and *BCL6* mutations (Fig. 3), being especially evident in the case of *MYC* and *PAX5*, where two distinct promoters correspond to separate mutation clusters (Fig. 1b, d). Specific features of the somatic hypermutation process, including the predominance of single nucleotide substitutions with occasional deletions or duplications, a preference for transitions over transversions and a specific motif targeting (RGYW: R = A/G, G, Y = C/T, W = A/T)²⁵, were also clearly recognizable in each of the four hypermutated loci (Table 1). The analysis of the nucleotide-exchange pattern showed an elevated ratio of G+C over A+T substitutions as in human V-region genes²⁶,

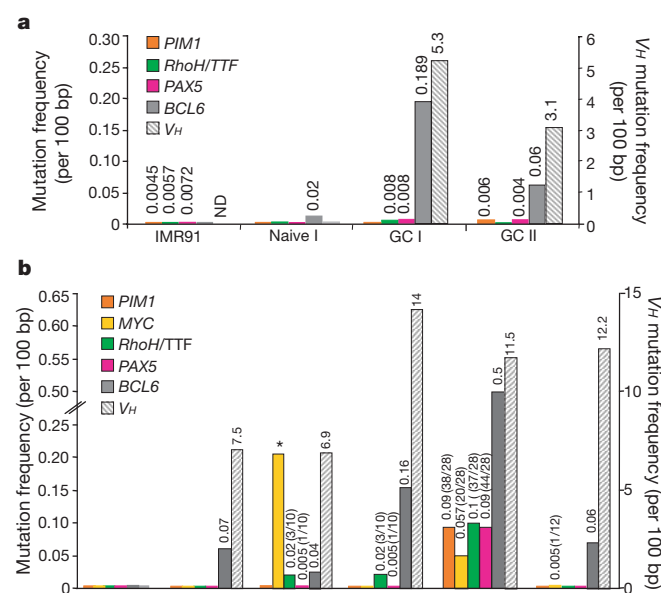


Figure 2 Mutation frequencies of *PIM1*, *MYC*, *RhoH/TTF*, *PAX5*, *BCL6* and immunoglobulin *V_H* in normal and transformed B cells. The exact value, when >0, is indicated above the bar. Note that a different scale (right) is used for the more highly mutated *V_H* sequences. **a**, Normal B-cell subpopulations, including naive (one donor) and germinal-centre (GC, two donors) B cells. The *V_H* data of donor 1 are from ref. 24. IMR91 is a fibroblast cell line used as a control. ND, not determined. **b**, B-cell malignancies. Ten cases each of the major histologic lymphoma subtypes (pre-germinal-centre mantle-cell lymphoma (MCL); germinal-centre-derived Burkitt (BL) and follicular (FL) lymphomas; post-germinal-centre multiple myeloma (MM); *V_H*-mutated CLL³⁰) were randomly selected along with 28 DLCL. From each patient, a fragment of *PIM1* (715 bp), *MYC* (832 bp) corresponding to the exon 2 area), *RhoH/TTF* (875 bp), *PAX5* (932 bp), *BCL6* (781 bp) and the rearranged *V_H* genes (~250 bp) were amplified and directly sequenced. Note that the mutation frequencies depicted here took into account both mutated and unmutated DLCL cases and only a subregion of the area affected by mutations (therefore the values are slightly different from those reported in Table 1, where only the mutated cases are counted). For *PIM1*, *MYC*, *PAX5* and *RhoH/TTF*, the total number of mutations per total number of cases analysed is also given in parentheses. Asterisk, endemic Burkitt lymphoma cases carrying a t(8;14) translocation.

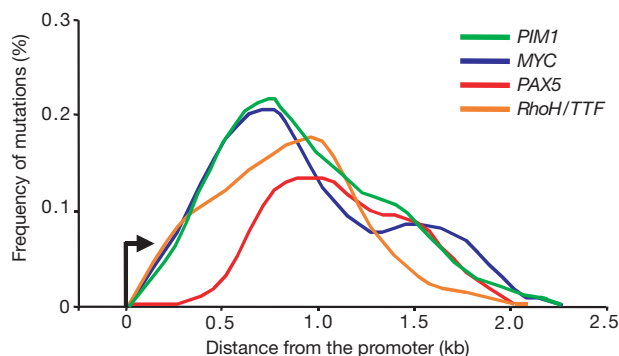


Figure 3 Distribution of somatic mutations from the transcription initiation site. Mutation frequency indicates the number of mutations per overlapping intervals of 500 bp in the mutated DLCL, and is plotted against the distance (kb) from the transcription initiation site (arrow).

whereas strand polarity could not be assessed owing to the small number of A+T changes (Supplementary Information Table 6). Overall, these observations are consistent with an abnormal activity of the same mechanism that generates *V_H*, *BCL6* and *Fas/CD95* mutations during normal B-cell development. Such aberrant function cannot be attributed to a higher activity of a normally operative somatic hypermutation process, as the same genes were not affected in follicular lymphoma or multiple myeloma, despite the fact that, in these tumours, *V_H* mutation frequencies were comparable to those detected in DLCL (Fig. 2b and ref. 3). A more plausible explanation is a malfunction of the somatic hypermutation mechanism in DLCL, leading to widespread alteration of genes that do not represent physiologic targets. This malfunction is unlikely to be due to defects in DNA mismatch repair, as DLCLs lack microsatellite instability²⁷, and seems not to involve the activation-induced deaminase (AID; a cytidine deaminase required for somatic hypermutation²⁸), because this gene was expressed and not mutated in 17 DLCL cases tested (not shown).

Notably, all of the genes affected by hypermutation (6 of 6, including *V_H* and *BCL6*) are also susceptible to chromosomal translocations in B-NHL. Furthermore, the genomic area targeted by mutations significantly overlaps with the major chromosomal breakpoint region within these genes (see Fig. 4). It has been shown that the process of somatic hypermutation is intrinsically associated with DNA double-strand breaks^{5,6} and it is speculated that this feature poses an intrinsic risk for the occurrence of chromosomal translocations⁴⁻⁶. Our results provide direct support for the model of hypermutation-associated chromosomal translocations, may explain the heterogeneity of translocations detected in DLCLs compared with other B-NHLs²⁹, and may account for translocations that cannot be attributed to mistakes in VDJ or switch recombination^{3,4}.

Aberrant targeting of somatic hypermutation represents a novel and powerful mechanism of malignant transformation. The relatively high frequency of mutated genes among those studied (4 of 17) suggests that the number of loci targeted in each DLCL case may be higher. Therefore, while chromosomal translocation may be one outcome of aberrant hypermutation in some loci, much more powerful appears the direct mutagenic effect on many genes, with consequences in part analogous to defects in DNA mismatch repair in colon carcinogenesis¹. As such, aberrant hypermutation of regulatory and coding sequences of multiple genes may provide the basis for DLCL pathogenesis and explain its phenotypic and clinical heterogeneity. □

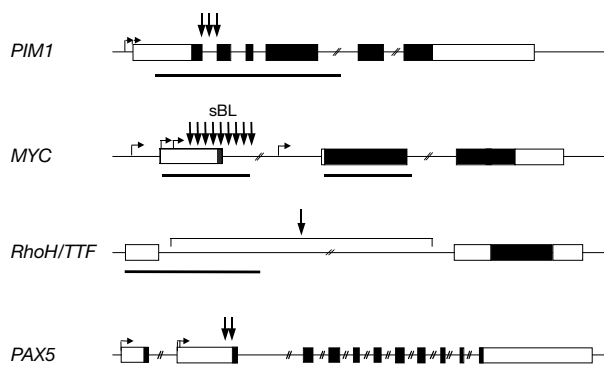


Figure 4 Hypermutable regions are susceptible to chromosomal translocations. Schematic diagram of the *PIM1*, *MYC*, *RhoH/TTF* and *PAX5* loci. A bar below the map indicates the region targeted by mutations in DLCL; large arrows represent translocation breakpoints in reported cases. For the *MYC* gene, numerous breakpoints have been mapped and only representative ones are shown. sBL, sporadic Burkitt lymphoma. The exact locations of the breakpoints within the *RhoH/TTF* intron 1 have not been reported.

Methods

Genomic sequences

The nucleotide sequences of *RhoH/TTF* (intron 1, partial) and *PAX5* (intron 1A and 1B) were obtained by PCR amplification and direct sequencing of genomic DNA from two healthy donors, and were submitted to GenBank under accession numbers AF386789, AF386790 and AF386791. The genomic sequence of *PIM1* differs from that reported in the database by one nucleotide and three 1-bp insertions, and has been deposited in GenBank under accession number AF386792.

Preparation and amplification of tumour DNA

Genomic DNA from 111 NHL cases, including 35 DLCL, 10 mantle-cell lymphoma, 13 B-cell CLL, 19 follicular lymphoma, 20 Burkitt lymphoma and 14 multiple myeloma, and from 4 DLCL cell lines, was extracted according to standard methods. The fraction of neoplastic cells corresponded to >70% in NHL and to ≈30% in multiple myeloma. Oligonucleotide sequences and PCR conditions for amplification of all of the genes analysed (0.5–3 kb from the transcription initiation site) are available on request.

Single cell analysis of normal B cells

Single naive and germinal-centre B cells were sorted from the reactive tonsils of two normal individuals based on the expression of specific markers as described^{4,7} and subjected to primer extension preamplification (PEP)⁹. A 4-μl PEP aliquot was then used as a template in a semi-nested PCR procedure in separate reactions for the amplification of *RhoH/TTF* (875 bp), *PAX5* (410-bp fragment, found mutated in 75% of the affected cases and containing 43% of the mutations), *BCL6* and the rearranged *V_H* genes. The protocol for amplification and sequencing of *BCL6* and *V_H* is as described⁷.

Cloning procedure

Genomic DNA was extracted from naive (IgD⁺CD27⁻) and germinal-centre (CD38⁺CD77⁺) B-cell populations purified from tonsillar tissue of two additional donors²⁴, and used for amplification of *PIM1* (715 bp), *RhoH/TTF* (875 bp), *PAX5* (932 bp) *BCL6* (781 bp) and the rearranged *V_H* genes using Pfu Turbo polymerase (Stratagene). The IMR91 fibroblast cell line was included as a control for background error rate. After 30 cycles (35 for *V_H*), PCR products were purified, incubated with 0.2 mM dATP and *Taq* polymerase (GIBCO BRL) for 15 min at 72 °C, and cloned into the pGEM-T vector (Promega) as recommended by the manufacturer. For each PCR product, 20–40 clones were analysed.

Sequencing analysis

Purified amplicons were sequenced directly from both strands as described⁷ and compared to the corresponding germline sequences; the first nucleotide of the reference sequence was arbitrarily defined as position +1. The presence of the mutations was always confirmed on independent PCR amplicons. Nucleotide changes due to previously reported polymorphisms or present in normal DNA from the same patient were disregarded in the assessment of the mutation frequency; in addition, all changes occurring more than once in separate cases were considered as polymorphic variants, unless their somatic origin could be proven.

Received 12 April; accepted 1 June 2001.

- Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643–649 (1998).
- Dalla-Favera, R. & Gaidano, G. in *Cancer. Principles and Practice of Oncology* (ed. DeVita, V. T. Jr, Hellman S. & Rosenberg, S. A.) 2215–2235 (Lippincott Williams & Wilkins, Philadelphia, 2001).
- Klein, U. *et al.* Somatic hypermutation in normal and transformed human B cells. *Immunol. Rev.* **162**, 261–280 (1998).
- Goossens, T., Klein, U. & Küppers, R. Frequent occurrence of deletions and duplications during somatic hypermutation: implications for oncogene translocations and heavy chain disease. *Proc. Natl Acad. Sci. USA* **95**, 2463–2468 (1998).
- Papavasiliou, F. N. & Schatz, D. G. Cell-cycle-regulated DNA double-stranded breaks in somatic hypermutation of immunoglobulin genes. *Nature* **408**, 216–221 (2000).
- Bross, L. *et al.* DNA double-strand breaks in immunoglobulin genes undergoing somatic hypermutation. *Immunity* **13**, 589–597 (2000).
- Pasqualucci, L. *et al.* BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. *Proc. Natl Acad. Sci. USA* **95**, 11816–11821 (1998).
- Shen, H. M., Peters, A., Baron, B., Zhu, X. & Storb, U. Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science* **280**, 1750–1752 (1998).
- Müschen, M. *et al.* Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction. *J. Exp. Med.* **192**, 1833–1840 (2000).
- Storb, U. *et al.* Cis-acting sequences that affect somatic hypermutation of Ig genes. *Immunol. Rev.* **162**, 153–160 (1998).
- Migliazza, A. *et al.* Frequent somatic hypermutation of the 5' noncoding region of the BCL6 gene in B-cell lymphoma. *Proc. Natl Acad. Sci. USA* **92**, 12520–12524 (1995).
- Rabbitts, T. H., Hamlyn, P. H. & Baer, R. Altered nucleotide sequences of a translocated c-myc gene in Burkitt lymphoma. *Nature* **306**, 760–765 (1983).
- Cuyper, H. T. *et al.* Murine leukemia virus-induced T-cell lymphomagenesis: integration of proviruses in a distinct chromosomal region. *Cell* **37**, 141–150 (1984).
- Akasaka, H. *et al.* Molecular anatomy of BCL6 translocations revealed by long-distance polymerase chain reaction-based assays. *Cancer Res.* **60**, 2335–2341 (2000).
- Hoover, D., Friedmann, M., Reeves, R. & Magnuson, N. S. Recombinant human pim-1 protein exhibits serine/threonine kinase activity. *J. Biol. Chem.* **266**, 14018–14023 (1991).
- Shirogane, T. *et al.* Synergistic roles for Pim-1 and c-Myc in STAT3-mediated cell cycle progression and antiapoptosis. *Immunity* **11**, 709–719 (1999).

17. Dang, C. V. c-Myc target genes involved in cell growth, apoptosis, and metabolism. *Mol. Cell. Biol.* **19**, 1–11 (1999).
18. Raffeld, M. *et al.* Clustered mutations in the transcriptional activation domain of Myc in 8q24 translocated lymphomas and their functional consequences. *Curr. Top. Microbiol. Immunol.* **194**, 265–272 (1995).
19. Cesarman, E., Dalla-Favera, R., Bentley, D. & Groudine, M. Mutations in the first exon are associated with altered transcription of c-myc in Burkitt lymphoma. *Science* **238**, 1272–1275 (1987).
20. Gu, W., Bhatia, K., Magrath, I. T., Dang, C. V. & Dalla-Favera, R. Binding and suppression of the Myc transcriptional activation domain by p107. *Science* **264**, 251–254 (1994).
21. Preudhomme, C. *et al.* Nonrandom 4p13 rearrangements of the RhoH/TTF gene, encoding a GTP-binding protein, in non-Hodgkin's lymphoma and multiple myeloma. *Oncogene* **19**, 2023–2032 (2000).
22. Morrison, A. M., Nutt, S. L., Thevenin, C., Rolink, A. & Busslinger, M. Loss- and gain-of-function mutations reveal an important role of BSAP (Pax-5) at the start and end of B cell differentiation. *Semin. Immunol.* **10**, 133–142 (1998).
23. Shen, H. M., Michael, N., Kim, N. & Storb, U. The TATA binding protein, c-Myc and survivin genes are not somatically hypermutated, while Ig and BCL6 genes are hypermutated in human memory B cells. *Int. Immunol.* **12**, 1085–1093 (2000).
24. Jungnickel, B. *et al.* Clonal deleterious mutations in the IκBα gene in the malignant cells in Hodgkin's lymphoma. *J. Exp. Med.* **191**, 395–402 (2000).
25. Neuberger, M. S. *et al.* Monitoring and interpreting the intrinsic features of somatic hypermutation. *Immunol. Rev.* **162**, 107–116 (1998).
26. Zan, H. *et al.* Induction of Ig somatic hypermutation and class switching in a human monoclonal IgM⁺ IgD⁺ B cell line in vitro: definition of the requirements and modalities of hypermutation. *J. Immunol.* **162**, 3437–3447 (1999).
27. Gambieri, B. *et al.* Microsatellite instability is rare in B-cell non-Hodgkin's lymphomas. *Blood* **89**, 975–979 (1997).
28. Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–563 (2000).
29. Cigudosa, J. C. *et al.* Cytogenetic analysis of 363 consecutively ascertained diffuse large B-cell lymphomas. *Genes Chromosom. Cancer* **25**, 123–133 (1999).
30. Pasqualucci, L., Neri, A., Baldini, L., Dalla-Favera, R. & Migliazza, A. BCL-6 mutations are associated with immunoglobulin variable heavy chain mutations in B-cell chronic lymphocytic leukemia. *Cancer Res.* **60**, 5644–5648 (2000).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank V. Miljkovic and A.-M. Babiac for assistance in DNA sequencing; C. Göttlinger for help with the single-cell sorting; M. Introna for providing the sequence of the *A-MYB* gene, and B. Jungnickel for providing DNA from sorted human naive and germinal-centre B-cell populations. We are grateful to U. Klein for discussions and to R. Baer for critically reading the manuscript. L.P. was a Fellow of the American Italian Cancer Foundation. P.N. was supported by the Max Kade Foundation. This work was supported by grants from the National Institute of Health to R.D.-F. and R.S.K.C., and from the Deutsche Forschungsgemeinschaft and a Heisenberg Award to R.K.

Correspondence and requests for materials should be addressed to R.D.-F. (e-mail: rd10@columbia.edu).

TAK1 is a ubiquitin-dependent kinase of MKK and IKK

Chen Wang^{*†}, Li Deng^{*†}, Mei Hong^{*}, Giridhar R. Akkaraju^{*}, Jun-ichiro Inoue[‡] & Zhijian J. Chen^{*}

^{*} Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9148, USA

[‡] Department of Applied Chemistry, Keio University, 3-14-1 Hiyoshi, Kouhoku-ku, Yokohama, Kanagawa 223-8522, Japan

[†] These authors contributed equally to this work

TRAF6 is a signal transducer that activates IκB kinase (IKK) and Jun amino-terminal kinase (JNK) in response to pro-inflammatory mediators such as interleukin-1 (IL-1) and lipopolysaccharides (LPS)^{1–4}. IKK activation by TRAF6 requires two intermediary factors, TRAF6-regulated IKK activator 1 (TRIKA1) and TRIKA2 (ref. 5). TRIKA1 is a dimeric ubiquitin-conjugating enzyme complex composed of Ubc13 and Uev1A (or the functionally equivalent Mms2). This Ubc complex, together with TRAF6, catalyses the formation of a Lys 63 (K63)-linked polyubiquitin

chain that mediates IKK activation through a unique proteasome-independent mechanism⁵. Here we report the purification and identification of TRIKA2, which is composed of TAK1, TAB1 and TAB2, a protein kinase complex previously implicated in IKK activation through an unknown mechanism^{6,7}. We find that the TAK1 kinase complex phosphorylates and activates IKK in a manner that depends on TRAF6 and Ubc13–Uev1A. Moreover, the activity of TAK1 to phosphorylate MKK6, which activates the JNK–p38 kinase pathway, is directly regulated by K63-linked polyubiquitination. We also provide evidence that TRAF6 is conjugated by the K63 polyubiquitin chains. These results indicate that ubiquitination has an important regulatory role in stress response pathways, including those of IKK and JNK.

We undertook seven chromatographic steps to purify TRIKA2 from HeLa cell cytoplasmic extracts (see Methods and Supplementary Information Fig. 1). To facilitate the identification of TRIKA2, we carried out immunoblotting analysis at each purification step using antibodies against proteins previously implicated in the IL-1 pathway⁸. Among the proteins tested—which include p62 (ref. 9), IRAK¹⁰, MEKK1 (ref. 11), MEKK3 (ref. 12), NIK¹³, RIP¹⁴ and TAK1 (ref. 6)—the only protein that co-purified with the TRIKA2 activity (stimulation of IKK in the presence of TRAF6 and Ubc13–Uev1A) in each of the purification steps was TAK1. As shown in Fig. 1a, fractions from the last MonoQ step that were capable of activating IKK contained TAK1 and its two binding partners, TAB1 and TAB2.

TAB2 normally resides in a membranous location in unstimulated cells, and it translocates to the cytosol on stimulation by IL-1 (ref. 7). The small amount of TAB2 found in the cytosolic extracts (S100) of unstimulated HeLa cells might have resulted from leakage owing to the hypotonic lysis protocol used for extraction of proteins¹⁵. To facilitate the purification of TRIKA2 (which might contain TAB2) we prepared HeLa cell extracts in a buffer containing 0.5% NP40 (see Methods). From these extracts, we used a TAB2-specific antibody to purify a complex containing stoichiometric amounts of TAK1, TAB1 and TAB2 as visualized by colloidal blue staining (Fig. 1b) and immunoblotting (data not shown). This immunopurified complex contained TRIKA2 activity as it stimulated IKK only in the presence of both Ubc13–Uev1A and TRAF6 (together with other ubiquitination components including E1 and ubiquitin).

To determine whether the TAK1–TAB1–TAB2 complex is indeed TRIKA2, we transfected expression vectors harbouring Flag-tagged TAK1 and TAB2 complementary DNA, respectively, into HEK-293 cells, and then immunopurified these proteins using a Flag-specific antibody. In each case, the transfected protein associated with endogenous partners to form a TAK1–TAB1–TAB2 complex, which activated IKK, but only in the presence of both Ubc13–Uev1A and TRAF6 (Fig. 1c). Notably, a point mutation in the ATP-binding domain of TAK1 (K63W) that abolished its kinase activity also abrogated its ability to stimulate IKK.

As the catalytic activity of TAK1 was essential for IKK activation, we examined whether TAK1 functioned as an IKK kinase (IKKK) to directly phosphorylate IKK at two specific serines (S177 and S181) in the activation loop¹⁶. To activate endogenous TAK1, we immunoprecipitated the TAK1 complex from HeLa cell extracts, which was then subjected to TRAF6-mediated ubiquitination reactions in the presence or absence of Ubc13–Uev1A. This TAK1 complex was then incubated with [γ -³²P]ATP together with recombinant IKKβ proteins, which include a catalytically inactive mutant (K44M) to prevent IKKβ autophosphorylation (Fig. 1d), and a triple-point mutant in which two serines in the activation loop are also mutated (K44M/S177A/S181A). Significantly, the TAK1 kinase once activated by Ubc13–Uev1A-mediated ubiquitination was able to phosphorylate IKKβ specifically at S177 and S181. Thus, TAK1 is a ubiquitin-dependent kinase of IKKβ.

To address the function of the individual components of the TAK1–TAB1–TAB2 complex in IKK activation, we expressed

TAK1, TAK1-TAB1, TAK1-TAB2 or coexpressed all three proteins in insect cells (Sf9) using the baculovirus expression system. These proteins were purified and then tested for IKK activation under the condition where the amount of TAK1 was equivalent (Fig. 2a). Only TAK1-TAB2 and TAK1-TAB1-TAB2 were able to stimulate IKK in the presence of Ubc13-Uev1A and TRAF6, suggesting that TAK1-TAB2 is the minimal sub-complex capable of activating IKK.

We carried out the above experiments using an IKK complex partially purified from HeLa cell extracts (greater than 60% pure). To determine whether the TAK1 complex could activate IKK in a reconstituted system using highly purified proteins, we further purified the IKK complex to homogeneity by immunoprecipitation using a Nemo-specific antibody (see Supplementary Information). The purified IKK complex contained exclusively IKK α , IKK β and Nemo, based on silver staining (Fig. 2b, lane 2) and immunoblotting (data not shown). We also obtained highly purified recombinant TAK1 complex (from Sf9 cells; Fig. 2b, lane 4) and 35 S-labelled I κ B α (see Methods). Using these purified proteins, together with purified E1, E2 (Ubc13-Uev1A) and E3 (TRAF6), we were able to reconstitute IKK activation by TRAF6 in a manner that was dependent on Ubc13-Uev1A (TRIKA1) and TAK1-TAB1-TAB2 (TRIKA2).

To determine whether the TAK1 complex is essential for IKK activation by TRAF6, we immunodepleted this complex from HEK-293 cell extracts using a TAB2 antibody (Fig. 2c). Removal of TAB2 from cell extracts abolished IKK activation by TRAF6, whereas immunoprecipitation with a control immunoglobulin- γ (IgG) antibody had no effect. Furthermore, addition of recombinant TAK1-TAB2 back into the depleted cell extracts restored IKK activation.

TRAF6 is essential for both IKK and JNK activation¹⁻⁴, and TAK1 can phosphorylate MKK6 (ref. 6), which in turn activates the JNK-p38 kinase pathway^{17,18}. Our finding that IKK is activated by TAK1 in a TRAF6- and Ubc13-Uev1A-dependent manner raises the possibility that the activation of MKK6 and subsequent activation of the JNK pathway might also be dependent on ubiquitin activation of TAK1. Indeed, when the endogenous TAK1 complex was subjected to ubiquitination by TRAF6 and Ubc13-Uev1A, it was activated to phosphorylate MKK6 (Fig. 3a). The ubiquitin-activated TAK1 kinase specifically phosphorylates MKK6 at S207 and T211 in the activation loop^{17,18} (Fig. 3b), allowing MKK6 to stimulate the kinase activity of JNK, which phosphorylates c-Jun at S63 and S73 (Fig. 3c). In further support of the role of ubiquitin in MKK/JNK activation, addition of methylated ubiquitin—which blocks

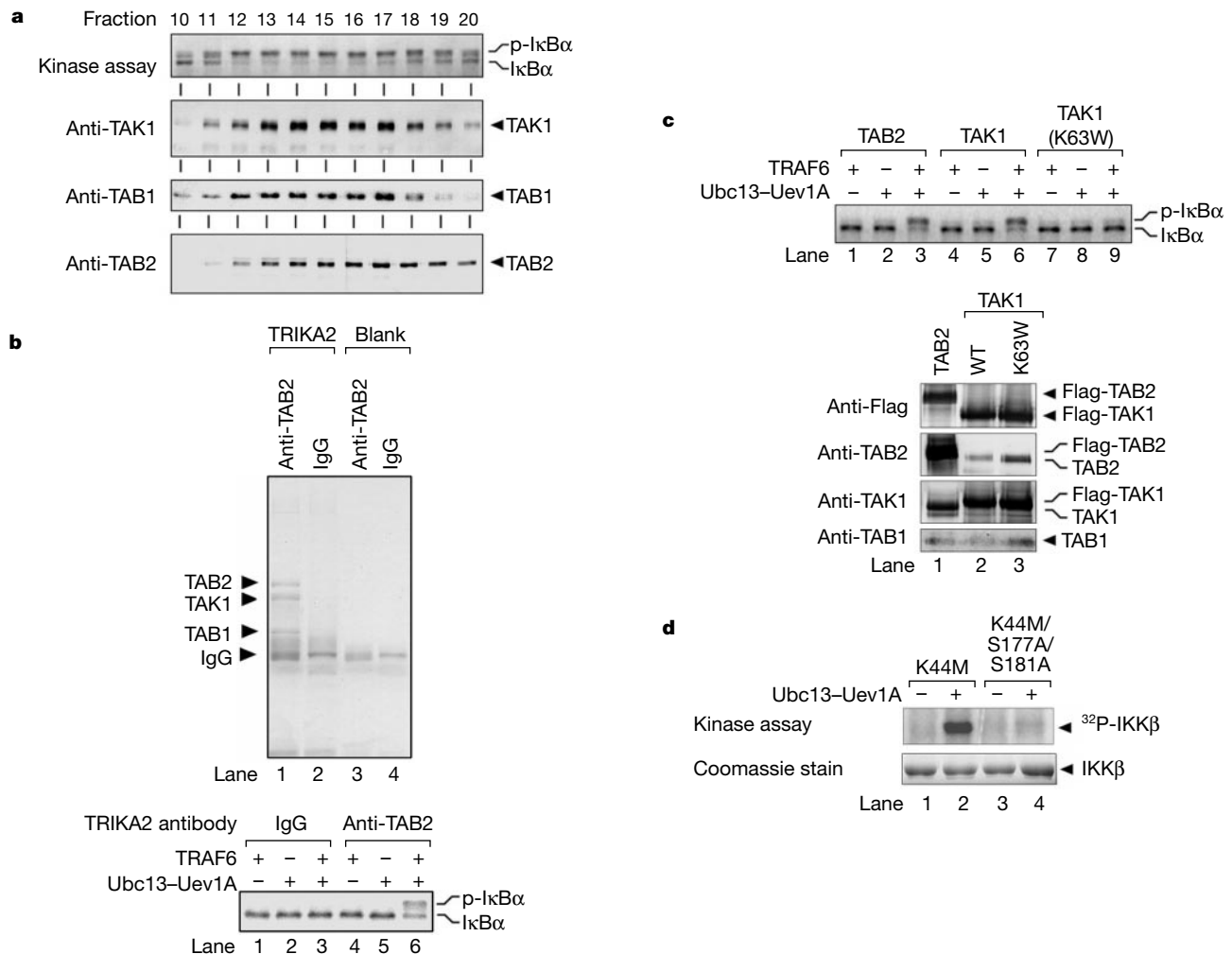


Figure 1 Identification of TRIKA2 as the TAK1-TAB1-TAB2 complex. **a**, Fractions from the last step of purification (MonoQ) were assayed for stimulation of IKK in the presence of Ubc13-Uev1A and TRAF6 together with other ubiquitination components. These fractions were immunoblotted as indicated. **b**, The TRIKA2 complex was purified using an immobilized TAB2 antibody (or IgG as control), and its composition analysed by colloidal blue staining (top). An aliquot of the immunopurified complex was assayed for IKK

activation (bottom). **c**, Expression vectors harbouring Flag-tagged TAK1, TAK1 (K63W) or TAB2 were transfected into HEK-293 cells, and the expressed proteins immunopurified for IKK activation assays (top) and immunoblotting (bottom). WT, wild type. **d**, Immunopurified TAK1 complex was activated by ubiquitination, and then assayed for its ability to phosphorylate recombinant IKK β (K44M) or IKK β (K44M/S177A/S181A) in the presence of [γ - 32 P]ATP.

polyubiquitination¹⁹—to cell extracts abolished the activation of both JNK and IKK by TRAF6 (Fig. 3d). Thus, ubiquitin activation of TAK1 provides a unifying mechanism for coordinate activation of the IKK and JNK pathways.

TRAF6 and Ubc13–Uev1A catalyse the synthesis of a unique polyubiquitin chain linked through K63 of ubiquitin, and the formation of such a chain is essential for IKK activation through a mechanism that does not involve proteasomal degradation⁵. To determine whether the formation of K63 chains is also important for TAK1 activation, we subjected the immunopurified TAK1 complex to ubiquitination reactions in the presence of various ubiquitin mutants (Fig. 3e). Notably, a point mutation at position 63 from lysine to arginine (K63R) completely abolished the ability of ubiquitin to stimulate TAK1 activity (Fig. 3e). Conversely, restoration of a single lysine at position 63 (K63) on an otherwise zero lysine background rescued the ability of ubiquitin to activate TAK1. Thus, a lysine at position 63 is both necessary and sufficient for ubiquitin to activate TAK1, most probably through the synthesis of K63-linked polyubiquitin chains.

Next, we sought to identify potential targets of ubiquitination, which could include components of the TAK1 complex and TRAF6 itself. To test whether these proteins are potential ubiquitination

targets, we treated HeLa cells with IL-1 β and then collected the cell extracts for immunoblotting with antibodies specific for each of these proteins. Stimulation of cells with IL-1 β led to the accumulation of TRAF6 conjugates with high molecular weights and with kinetics that correlated with the degradation and re-synthesis of I κ B α (Fig. 4a). We confirmed that these conjugates were polyubiquitinated TRAF6 by immunoprecipitating TRAF6 followed by immunoblotting with a ubiquitin-specific antibody. No detectable ubiquitination of components of the TAK1 complex was observed in the same experiment.

It has been suggested that oligomerization of TRAF6 through its carboxy-terminal TRAF domain leads to the activation of the IKK and JNK pathways²⁰. To determine whether oligomerization of TRAF6 triggers its ubiquitination, we sought to create a cell line in which the activation of IKK is controlled by the inducible oligomerization of TRAF6. We replaced the C-terminal TRAF domain of TRAF6 with a fragment of the bacterial gyrase B, which dimerizes in the presence of coumermycin A (ref. 21) (Fig. 4b). The chimaeric construct (T6RZC) was transfected into HEK-293 cells to establish a stable line in which I κ B α was rapidly degraded on the addition of coumermycin A (Fig. 4c). Concomitant with the degradation of I κ B α , a ladder of polyubiquitinated T6RZC

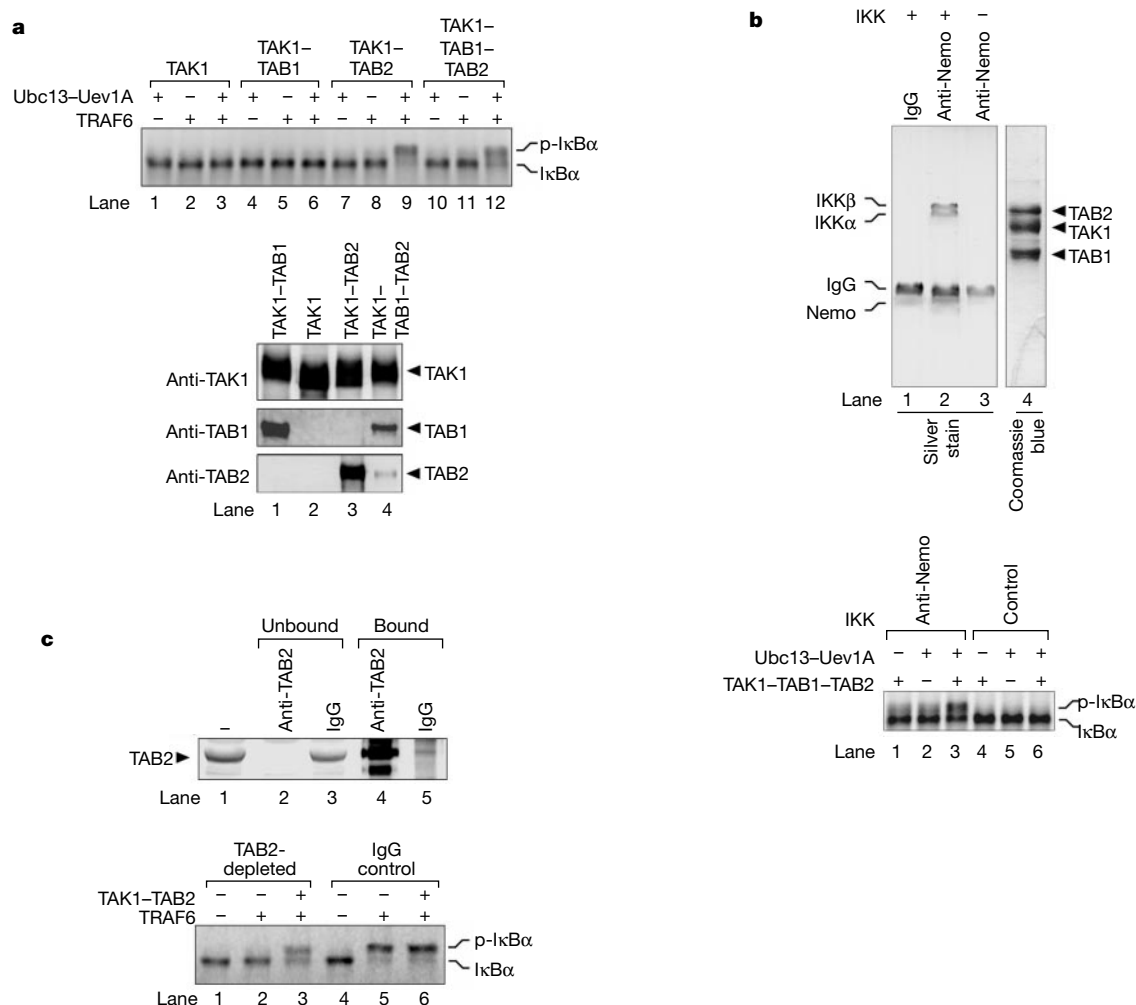


Figure 2 The TAK1 complex is both sufficient and necessary to activate IKK in conjunction with TRAF6 and Ubc13–Uev1A. **a**, Recombinant TAK1, TAK1–TAB1, TAK1–TAB2 and TAK1–TAB1–TAB2 were expressed in insect cells (Sf9) and purified. These protein complexes containing an equivalent amount of TAK1 (bottom) were assayed for IKK activation (top). **b**, IKK complex was purified using a Nemo antibody (or IgG as control, top) and then tested for its kinase activity (bottom) in a reconstituted system containing purified

E1, ubiquitin, Ubc13–Uev1A (E2), TRAF6 (E3), TAK1–TAB1–TAB2 and ³⁵S-labelled I κ B α . Lane 3 (top) is the background staining of the Nemo antibody. **c**, HEK-293 cell extracts were immunodepleted of TAB2 (top) and then assayed for IKK activation (bottom). In lanes 3 and 6, recombinant TAK1–TAB2 was added back to the assays. Unbound, supernatant after immunoprecipitation; bound, immunoprecipitated beads.

accumulated (Fig. 4c, d). Notably, although treatment of cells with coumermycin A (Fig. 4c) or IL-1 β (Fig. 4a) induced the rapid degradation of I κ B α , no evidence of T6RZC or TRAF6 degradation was observed during the same time course. Deletion of a coil-coiled region (residues 292–358) in TRAF6 (T6RZ) abolishes its ubiquitination as well as its ability to stimulate I κ B α degradation, suggesting a correlation between the ability of TRAF6 to be ubiquitinated and its ability to stimulate the IKK pathway. After ubiquitination, TRAF6 led to the activation of IKK in the presence of the TAK1 complex without further requirements for additional

ubiquitination enzymes such as E1 and Ubc13–Uev1A (Fig. 4e). The polyubiquitin chains on T6RZC were apparently K63-linked, as it was polyubiquitinated by ubiquitin mutants containing K63 but not R63 (Fig. 4f).

TAK1 was thought to activate IKK through the intermediary kinase NIK⁶; however, genetic studies have shown that NIK is not involved in IKK activation by most stimuli, including IL-1 β ²². Thus, it is unlikely that NIK is the link between TAK1 and IKK. Instead, our data strongly suggest that TAK1 is an IKKK that phosphorylates and activates IKK in the TRAF6 pathway. How TAK1 itself is

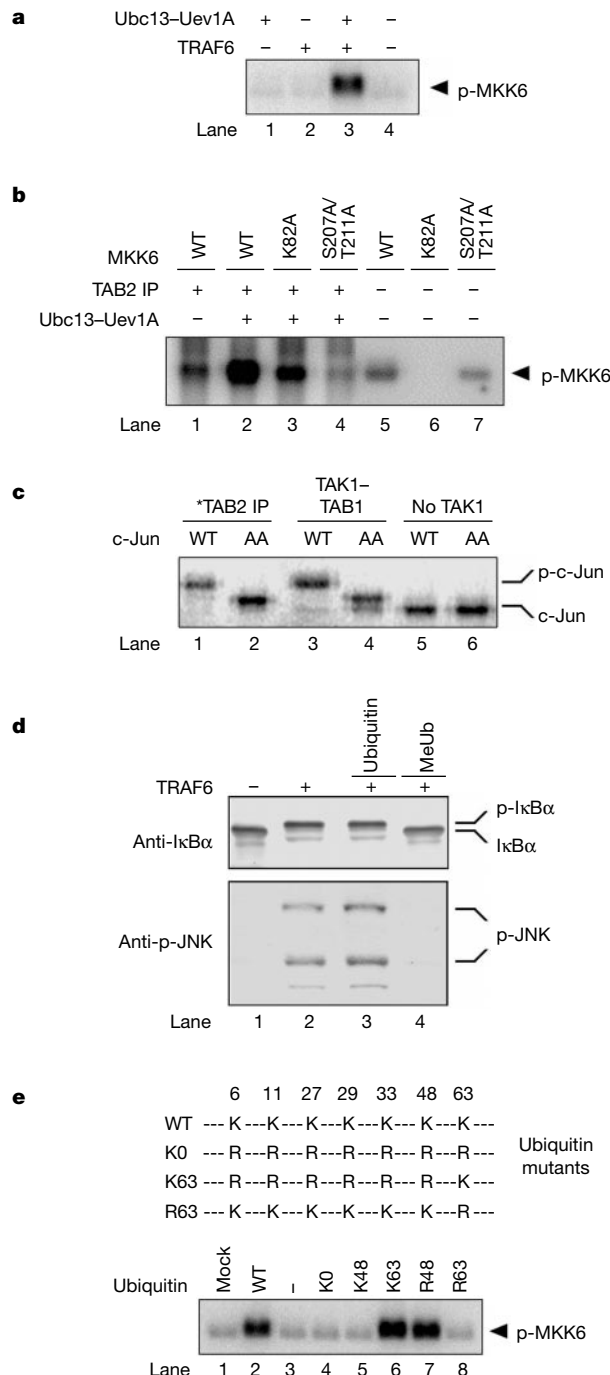


Figure 3 TAK1 is a ubiquitin-dependent kinase of the MKK–JNK pathway. **a**, **b**, The TAK1 complex was activated by ubiquitination and then assayed for phosphorylation of MKK6 or MKK6 mutants in the presence of [γ -³²P]ATP. WT, wild type; TAB2 IP, immunopurified TAK1 complex; K82A, catalytically inactive MKK6; S207A/T211A, mutations in the activation loop of MKK6. **c**, Ubiquitin-activated TAK1 complex (indicated with an asterisk) was assayed for JNK activation in a coupled reaction containing recombinant MKK6, JNK

and ³⁵S-labelled c-Jun or c-Jun mutant (S63A/S73A: labelled AA). **d**, 70Z cell extracts were incubated with TRAF6 in the presence of ubiquitin or methylated ubiquitin (MeUb) as described⁵, and then analysed by immunoblotting with antibodies against I κ B α or phospho-JNK. **e**, Immunopurified TAK1 complex was subjected to ubiquitination in the presence of Ubc13–Uev1A and TRAF6 together with various ubiquitin mutants, and then assayed for phosphorylation of MKK6. K0, ubiquitin mutant containing no lysine.

activated was previously unknown. Although recombinant TAK1 can be activated by TAB1, endogenous TAK1 is inactive even though it is always associated with TAB1 and TAB2 (refs 6, 7). Thus, association of TAB1 with TAK1 *per se* cannot account for the activation of the endogenous TAK1 complex. Our results indicating that TRAF6 is ubiquitinated, and that the formation of K63-linked polyubiquitin chains directly activates the endogenous TAK1 complex in a biochemically well defined system provide a new mechanism for the activation of TAK1. According to this mechanism, stimulation of cells with IL-1 β leads to the oligomerization of TRAF6, which triggers its ubiquitination through the action of Ubc13–Uev1A. IL-1 β treatment also triggers the release of TAB2 from a membraneous location to cytoplasm⁷, where it binds to

ubiquitinated TRAF6 as well as TAK1. Exactly how TAK1 is activated by ubiquitinated TRAF6 is unclear. One possibility is that polyubiquitination may convert the TRAF6–TAB2 complex into an activator of TAK1 (for example, through a conformational change). Alternatively, the K63-linked polyubiquitin chains attached to TRAF6 may reach into TAK1 to activate it as a result of complex formation among TRAF6, TAB2 and TAK1. It is also formally possible that there is a transient but undetectable level of polyubiquitination on TAK1, and that this dynamic ubiquitination of TAK1 leads directly to its activation. Notwithstanding this uncertainty concerning the detailed mechanism of TAK1 activation, our results demonstrate a crucial role of K63-linked polyubiquitination in TAK1 activation.

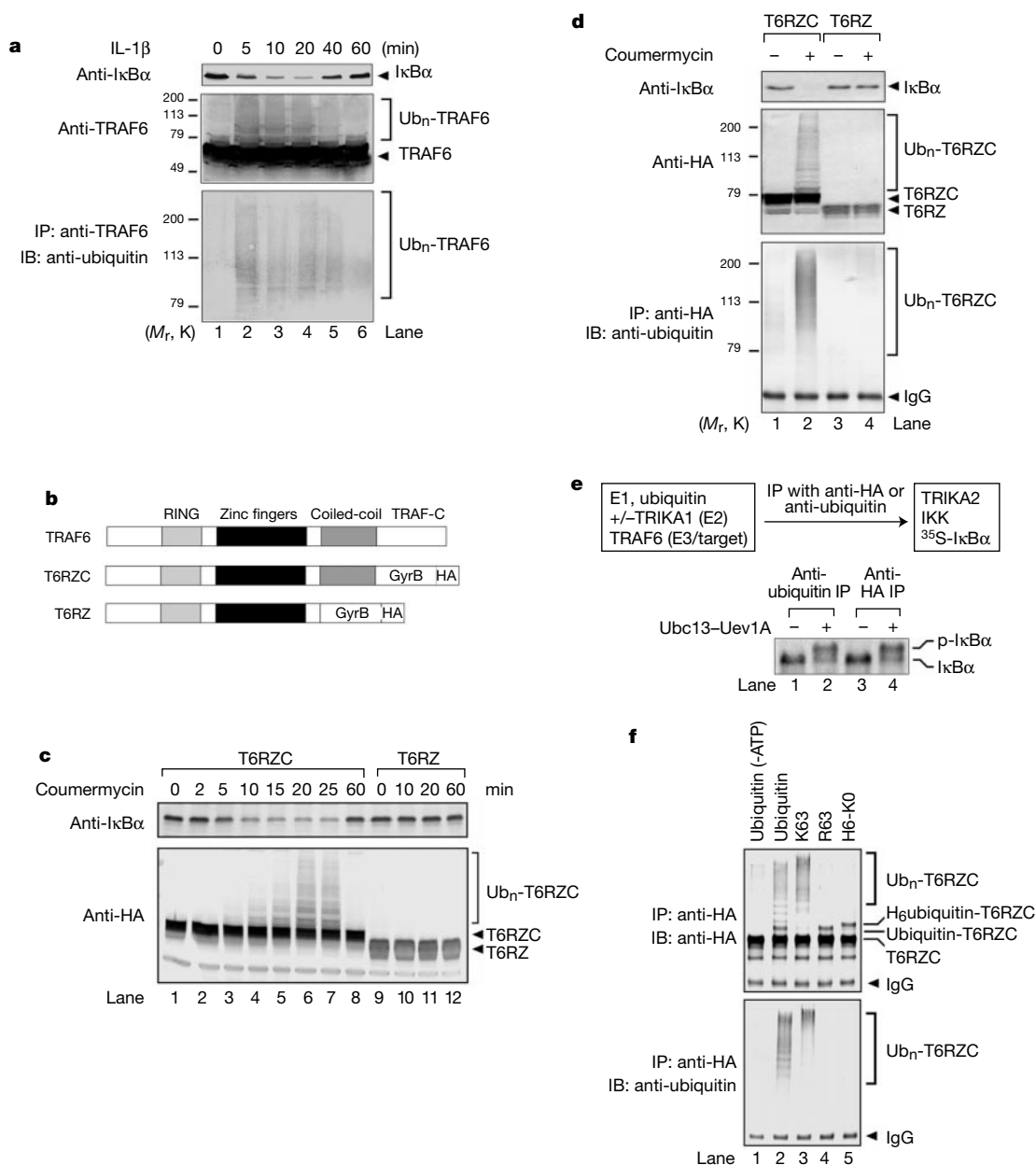


Figure 4 K63-linked polyubiquitination of TRAF6. **a**, Extracts from IL-1 β -stimulated HeLa cells were analysed by immunoblotting with antibodies against I κ B α (top) or TRAF6 (middle). These extracts were also immunoprecipitated with the TRAF6 antibody followed by immunoblotting with a ubiquitin antibody (bottom). **b**, Diagrams of TRAF6–gyrase B chimaeric constructs. **c**, HEK-293 T6RZC or HEK-293 T6RZ cells were treated with coumermycin A, and cell extracts were collected for immunoblotting. **d**, T6RZC and T6RZ were immunoprecipitated from coumermycin-stimulated cells with an HA antibody and

analysed by immunoblotting with a ubiquitin antibody. **e**, Ubiquitination reactions were carried out as shown and ubiquitin conjugates immunoprecipitated with antibodies against HA (for TRAF6) or ubiquitin. The beads were assayed for stimulation of IKK in the presence of TAK1 complex (TRIKA2). **f**, T6RZC was ubiquitinated in the presence of various ubiquitin mutants, and the conjugates immunoprecipitated with an HA antibody and then analysed by immunoblotting. H₆-KO, His₆-tagged ubiquitin mutant containing no lysine; ub_n, polyubiquitin.

Ubiquitin-dependent activation of TAK1 offers a possible solution to a conceptual problem inherent in kinase cascades: if kinase C is activated by kinase B, which is in turn activated by kinase A, how is kinase A initially activated? By definition, the first kinase cannot be activated by another kinase, so the initial activation has to be a kinase-independent mechanism. The formation of K63-linked polyubiquitin chains may provide such a mechanism, at least in the case of TAK1 activation. Activated TAK1 then phosphorylates IKK and MKK6, leading to the activation of NF- κ B and JNK-p38 kinase pathways, respectively. These results raise the possibility that polyubiquitination through K63 of ubiquitin may be a general mechanism for the regulation of stress kinase pathways, including those of IKK and JNK. □

Methods

Reagents

We obtained HeLa S3 cells from the National Culture Center (Minneapolis). The TAB2 antibody was produced from rabbits by immunization using a His₆-TAB2 N-terminal fragment (residues 1–450) as an antigen, and then affinity purified on a TAB2 column. Antibodies against TAK1, TAB1, ubiquitin and I κ B α were obtained from Santa Cruz Biotech. We obtained the antibody against phospho-JNK from New England Biolabs. cDNAs encoding TAK1, TAB1, TAB2 and MKK6 were cloned by PCR from cDNA libraries based on published sequences^{6,7,17} and verified by sequencing. pcDNA3-JNK, -Jun and -c-Jun (S63A/S73A) were provided by F. Lee¹¹. cDNAs encoding TRAF6, Ubc13, Uev1A and I κ B α have been previously described⁵. *In vitro* translation of I κ B α and c-Jun were carried out using wheat germ extracts supplemented with ³⁵S-labelled methionine²³. ³⁵S-labelled I κ B α , which has a Flag tag at its N terminus, was further purified by immunoaffinity chromatography. We carried out site-directed mutagenesis using the QuikChange kit (Stratagene). Recombinant MKK6 and JNK were expressed in *Escherichia coli* as His₆-tagged proteins and purified using nickel columns. TAK1, TAB1 and TAB2 were expressed in insect (Sf9) cells individually or in combination after infection or co-infection of baculovirus harbouring the respective genes in pFAST-Bac (Gibco-BRL), and then purified by means of nickel affinity column followed by MonoQ or gel filtration chromatography. Similarly, recombinant IKK β (K44M), IKK β (K44M/S177A/S181) and Nemo were expressed in Sf9 cells, except that these proteins were engineered to contain a Flag epitope at their N termini and were purified to homogeneity by Flag immunoaffinity chromatography. Other reagents, including E1, TRIKA1 (E2), TRAF6 (E3) and various ubiquitin mutants have been described previously^{5,24}.

Purification of TRIKA2

Crude cytosolic extracts (S100) from 100-l culture of HeLa cells were prepared as described⁵, and applied to a Q-Sepharose column pre-equilibrated with buffer A (20 mM Tris-HCl, pH 7.5, 1 mM dithiothreitol (DTT), 0.5 mM phenylmethyl sulphonyl fluoride (PMSF), 1 μ g ml⁻¹ leupeptin) containing 0.1 M NaCl. Proteins bound to the column were eluted with buffer A plus 0.21 M NaCl, and then precipitated with ammonium sulphate (40%). The pellets were resuspended and loaded onto a Phenyl-Sepharose column. Bound proteins were eluted with a gradient of decreasing concentration of ammonium sulphate (0.5–0 M) in buffer A. The fractions containing TRIKA2 activity were applied to a Heparin-Sepharose column and eluted with a gradient of NaCl (0–0.3 M) in buffer A. Active TRIKA2 fractions were concentrated before loading onto a Superdex-200 column pre-equilibrated with buffer B (20 mM Tris-HCl, pH 7.5, 0.5 mM DTT, 0.2 mM PMSF, 0.15 M NaCl, 15% glycerol). Proteins eluted from the Superdex column were applied to a Mono-S column and eluted with a gradient of NaCl (0–0.3 M) in buffer C (20 mM HEPES, pH 7.0, 0.5 mM DTT, 0.2 mM PMSF, 0.02% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulphonate (CHAPS), 10% glycerol). Fractions containing TRIKA2 were diluted in buffer D (20 mM Tris-HCl, pH 7.5, 0.5 mM DTT, 0.2 mM PMSF, 0.02% CHAPS, 10% glycerol), loaded onto a Mono-Q column, and then eluted with a gradient of NaCl (0.1–0.3 M) in buffer D. For immunopurification of TRIKA2 (Fig. 1b), cell extracts were prepared as described for S100 except that HeLa cells were lysed in buffer A supplemented with 0.5% NP40 and 0.1 M NaCl. TRIKA2 was partially purified by conventional chromatographic steps including Q-Sepharose, ammonium sulphate precipitation, Heparin-Sepharose and MonoQ, as described above. The fractions containing TRIKA2 activity from MonoQ were further purified by immunoprecipitation using the TAB2 antibody immobilized on Protein A-Sepharose.

Kinase assays

IKK activity was measured in a reconstituted system (10 μ l) containing an ATP buffer (50 mM Tris-HCl, pH 7.5, 5 mM MgCl₂, 2 mM ATP), ³⁵S-labelled I κ B α (1 μ l), purified IKK complex (5 nM), E1 (50 nM), Ubc13–Uev1A (0.3 μ M), TRAF6 (0.1 μ M), ubiquitin (50 μ M) and various amounts of partially purified or immunopurified TRIKA2 (TAK1 complex). After incubation at 30 °C for 1 h, the reaction products were resolved by SDS–PAGE and analysed using a PhosphorImager.

To assay the activity of TAK1, the kinase complex immobilized on the anti-TAB2 beads was incubated with E1, Ubc13–Uev1A, TRAF6, ubiquitin and the ATP buffer at room temperature for 1 h under conditions similar to those used for IKK activation (see above). The beads were then washed extensively with TBS (20 mM Tris-HCl, pH 7.5, 150 mM NaCl) plus 0.5% NP40, followed by incubation with recombinant MKK6 (1 μ M),

[γ -³²P]ATP (0.5 μ Ci ml⁻¹), and a kinase buffer (50 mM Tris-HCl, pH 7.5, 0.5 mM DTT, 5 mM MgCl₂, 50 μ M ATP). Phosphorylation of IKK β was carried out similarly, except that IKK β was pre-incubated with Nemo to form a complex, which then served as a substrate for TAK1. The reactions were carried out at 30 °C for 1 h and then analysed by SDS–PAGE. For assays of JNK activity, the ubiquitin-activated TAK1 beads were incubated with MKK6 (60 nM), JNK (0.1 μ M), ³⁵S-labelled c-Jun or c-Jun (S63A/S73A, 1 μ l) and the ATP buffer. Phosphorylation of c-Jun was then determined based on its mobility shift on SDS–PAGE.

Generation of TRAF6–gyrase B stable cell lines

The TRAF6–gyrase B fusion was constructed by replacing the C-terminal TRAF domain of TRAF6 (residues 1–358) with the N-terminal fragment of bacterial gyrase B (T6RZC). Another chimaeric construct contained residues 1–291 of TRAF6 (T6RZ). These chimaeras were cloned into the mammalian expression vector pEF-IRES-P (ref. 5). After transfection of DNA into HEK-293 cells, stable clones were selected in puromycin (1 μ g ml⁻¹). To induce the dimerization of TRAF6 and subsequent activation of the JNK and IKK pathways, cells were treated with coumermycin A (1 μ M) for different lengths of time, and cell extracts were prepared for further analysis.

Received 2 May; accepted 29 May 2001.

1. Cao, Z. *et al.* TRAF6 is a signal transducer for interleukin-1. *Nature* **383**, 443–446 (1996).
2. Ishida, T. *et al.* Identification of TRAF6, a novel tumor necrosis factor receptor-associated factor protein that mediates signaling from an amino-terminal domain of the CD40 cytoplasmic region. *J. Biol. Chem.* **271**, 28745–28748 (1996).
3. Lomaga, M. A. *et al.* TRAF6 deficiency results in osteopetrosis and defective interleukin-1, CD40, and LPS signaling. *Genes Dev.* **13**, 1015–1024 (1999).
4. Naito, A. *et al.* Severe osteopetrosis, defective interleukin-1 signalling and lymph node organogenesis in TRAF6-deficient mice. *Genes Cells* **4**, 353–362 (1999).
5. Deng, L. *et al.* Activation of the I κ B kinase complex requires a dimeric ubiquitin conjugating enzyme complex and a unique polyubiquitin chain. *Cell* **103**, 351–361 (2000).
6. Nimomiya-Tsuji, J. *et al.* The kinase TAK1 can activate the NIK-I κ B as well as the MAP kinase cascade in the IL-1 signalling pathway. *Nature* **398**, 252–256 (1999).
7. Takaue, G. *et al.* TAB2, a novel adaptor protein, mediates activation of TAK1 MAPKKK by linking TAK1 to TRAF6 in the IL-1 signal transduction pathway. *Mol. Cell* **5**, 649–658 (2000).
8. Zhang, G. & Ghosh, S. Toll-like receptor-mediated NF- κ B activation: a phylogenetically conserved paradigm in innate immunity. *J. Clin. Invest.* **107**, 13–19 (2001).
9. Sanz, L. *et al.* The atypical PKC-interacting protein p62 channels NF- κ B activation by the IL-1-TRAF6 pathway. *EMBO J.* **19**, 1576–1586 (2000).
10. Cao, Z., Henzel, W. J. & Gao, X. IRAK: a kinase associated with the interleukin-1 receptor. *Science* **271**, 1128–1131 (1996).
11. Lee, F. S., Hagler, J., Chen, Z. & Maniatis, T. Activation of the I κ B kinase complex by MEKK1, a kinase of the JNK pathway. *Cell* **88**, 213–222 (1997).
12. Zhao, Q. & Lee, F. S. Mitogen-activated protein kinase/ERK kinase kinases 2 and 3 activate nuclear factor- κ B through I κ B kinase- α and I κ B kinase- β . *J. Biol. Chem.* **274**, 8355–8358 (1999).
13. Malinin, N. L. *et al.* MAP3K-related kinase involved in NF- κ B induction by TNF, CD95 and IL-1. *Nature* **385**, 540–544 (1997).
14. Ting, A. T. *et al.* RIP mediates tumor necrosis factor receptor 1 activation of NF- κ B but not Fas/APO-1-initiated apoptosis. *EMBO J.* **15**, 6189–6196 (1996).
15. Liu, X. *et al.* Induction of apoptotic program in cell-free extracts: requirement for dATP and cytochrome c. *Cell* **86**, 147–157 (1996).
16. Mercurio, F. *et al.* IKK-1 and IKK-2: cytokine-activated I κ B kinases essential for NF- κ B activation. *Science* **278**, 860–866 (1997).
17. Raingeaud, J. *et al.* MKK3- and MKK6-regulated gene expression is mediated by the p38 mitogen-activated protein kinase signal transduction pathway. *Mol. Cell Biol.* **16**, 1247–1255 (1996).
18. Davis, R. J. Signal transduction by the JNK group of MAP kinases. *Cell* **103**, 239–252 (2000).
19. Herskko, A. & Heller, H. Occurrence of a polyubiquitin structure in ubiquitin-protein conjugation. *Biochem. Biophys. Res. Commun.* **128**, 1079–1086 (1985).
20. Baud, V. *et al.* Signaling by proinflammatory cytokines: oligomerization of TRAF2 and TRAF6 is sufficient for JNK and IKK activation and target gene induction via an amino-terminal effector domain. *Genes Dev.* **13**, 1297–1308 (1999).
21. Farrar, M. A. *et al.* Activation of the Raf-1 kinase cascade by coumermycin-induced dimerization. *Nature* **383**, 178–181 (1996).
22. Yin, L., Defective lymphotoxin-beta receptor-induced NF- κ B transcriptional activity in NIK-deficient mice. *Science* **291**, 2162–2165 (2001).
23. Chen, Z. J., Parent, L. & Maniatis, T. Site-specific phosphorylation of I κ B α by a novel ubiquitination-dependent protein kinase activity. *Cell* **84**, 853–862 (1996).
24. Hofmann, R. M. & Pickart, C. M. Noncanonical MMS2-encoded ubiquitin-conjugating enzyme functions in assembly of novel polyubiquitin chains for DNA repair. *Cell* **96**, 645–653 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank C. Pickart for providing expression constructs encoding ubiquitin lysine mutants and F. Lee for providing cDNAs encoding JNK, c-Jun and c-Jun (S63A/S73A). We also thank E. Spencer, L. Yang and A. Braun for technical support, and A. Tizenor and S. Johnson for graphic work. We are grateful to E. Olson, T. Maniatis and J. Jiang for critically reading the manuscript. This work was supported in part by grants from the National Institute of Health and the Welch Foundation. Z.J.C. is a Searle Scholar supported by The Chicago Community Trust.

Correspondence and requests for materials should be addressed to Z.J.C. (e-mail: jchen@hamon.swmed.edu).

A MAP kinase-dependent actin checkpoint ensures proper spindle orientation in fission yeast

Yannick Gachet^{*†}, Sylvie Tournier^{‡†}, Jonathan B. A. Millar[‡] & Jeremy S. Hyams^{*}

^{*} Department of Biology, University College London, Gower Street, London WC1E 6B, UK

[‡] Division of Yeast Genetics, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

[†] These authors contributed equally to this work

The accurate segregation of chromosomes at mitosis depends on a correctly assembled bipolar spindle that exerts balanced forces on each sister chromatid^{1,2}. The integrity of mitotic chromosome segregation is ensured by the spindle assembly checkpoint (SAC) that delays mitosis in response to defective spindle organisation or failure of chromosome attachment^{2,3}. Here we describe a distinct mitotic checkpoint in the fission yeast, *Schizosaccharomyces pombe*, that monitors the integrity of the actin cytoskeleton and delays sister chromatid separation, spindle elongation and cytokinesis until spindle poles have been properly oriented. This mitotic delay is imposed by a stress-activated mitogen-activated protein (MAP) kinase pathway but is independent of the anaphase-promoting complex (APC)^{4,5}.

The actin cytoskeleton undergoes marked cell-cycle-dependent remodelling in all eukaryotes. In fission yeast, actin patches underlie the regions of cell extension at the cell tips during interphase. These disappear as cells enter mitosis and are replaced by a medial ring at the cell cortex that constricts during cytokinesis⁶. To investigate the coordination between the actin cytoskeleton and the fission yeast cell cycle, cells were synchronized in early G2 and re-inoculated into fresh medium containing latrunculin B (Lat B), an inhibitor of actin polymerization⁷. Whereas control cells underwent nuclear division, Lat B-treated cells remained uninucleate for a further 60 min (Fig. 1a). At 90 min after re-inoculation, control cells showed elongated mitotic spindles and balanced astral microtubules⁸, whereas Lat B-treated cells arrested in mitosis with up to 30% of the mitotic spindles being short (< 2.5 μ m) and frequently mis-oriented, with either missing or unbalanced astral microtubules (Fig. 1b).

The formation of bipolar spindles was not affected by Lat B treatment, indicating that cells were not delayed for mitotic entry (Fig. 1c). However, although the appearance of short spindles was transient in control cells^{9,10}, in the presence of Lat B spindle elongation was substantially delayed (Fig. 1d). In control cells the mean angle of displacement (α) of short spindles from the long axis of the cell was $12 \pm 0.9^\circ$, whereas for Lat B-treated cells displacement was increased to $27 \pm 1.3^\circ$ (s.e.m.; $n = 200$). This is an underestimate of the misorientation as spindle position in the z axis was not accounted for. A similar mitotic delay was observed in cells treated with latrunculin A and cytochalasin D, and in cultures of the temperature-sensitive actin mutant *cps8* (ref. 11), where 26% of the cell population displayed short, misoriented spindles at the permissive temperature. By contrast only 0.8% of wild-type cells displayed short mitotic spindles in a log phase population (Fig. 1e, f). We conclude that disruption of the actin cytoskeleton perturbs proper spindle orientation and imposes a cell-cycle delay in mitosis.

To assess whether Lat B delays sister chromatid separation, centromeres were visualized in a strain that expresses the LacI DNA-binding domain fused to green fluorescent protein (GFP) and contains repetitive units of the LacO sequence integrated near

to centromere I (ref. 10). At a time when synchronized control cells displayed two clearly separated centromeres at the ends of elongated mitotic spindles, a single signal was apparent on the short mitotic spindles of Lat B-treated cells (Fig. 2a). We next determined whether lack of centromere separation is due to inhibition of sister chromatid separation or of spindle elongation. Cohesion between sister chromatids is maintained by a cohesin complex that includes the fission yeast Rad21 protein^{12,13}. We found that a *rad21-45* mutant was unable to arrest in the presence of Lat B, indicating that depolymerization of the actin cytoskeleton primarily prevents sister chromatid separation rather than spindle elongation (Fig. 2b). Notably, synchronized *rad21-45* cells lost viability as they passed through mitosis only in the presence of Lat B and not in its absence (Fig. 2c). Cleavage of Rad21 by a specialized protease, Cut1 (Separase), is essential for sister chromatid separation in anaphase and can be monitored by the appearance of a Rad21 cleavage product¹³. To examine more closely the mechanism by which Lat B inhibits centromere separation, *cdc25-22* cells were synchronously released from a G2 arrest in the presence of Lat B (Fig. 2d). Whereas control cells rapidly entered mitosis, Lat B-treated cells arrested in mitosis with undivided nuclei, over 80% of which displayed condensed chromatin and short, misaligned mitotic spindles (Fig. 2d, e). Appearance of the Rad21 cleavage product was observed as control cells entered anaphase but was substantially delayed in Lat B-treated cultures (Fig. 2f). These results indicate that disruption of the actin cytoskeleton delays Cut1-mediated cohesin cleavage and, as a consequence, sister chromatid separation.

When the mitotic spindle is damaged or bipolar attachment of chromosomes to spindle microtubules is lost a metaphase delay

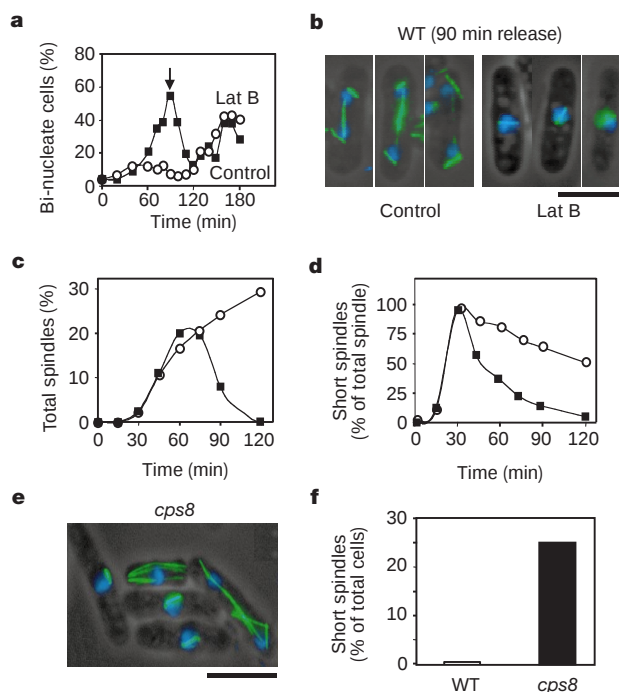


Figure 1 Latrunculin B induces a mitotic delay. **a**, Nuclear division in synchronized wild-type cells in the presence (circles) or absence (squares) of 10 μ M Lat B. **b**, Control and Lat B cells at 90 min (arrow in **a**) probed with anti-tubulin antibodies (green) and 4,6-diamidino-2-phenylindole (DAPI) (blue). Scale bar, 10 μ m. **c**, Percentage of cells with mitotic spindles in synchronized cells with (circles) or without (squares) 10 μ M Lat B. **d**, Percentage of mitotic spindles in **c** that measure less than 2.5 μ m. **e**, Log phase *cps8* mutant cells probed with anti-tubulin antibodies (green) and stained with DAPI (blue). Scale bar, 10 μ m. **f**, Percentage of cells in log phase wild-type (white bar) or *cps8* (black bar) populations with mitotic spindles measuring less than 2.5 μ m.

occurs. This delay is brought about by recruitment of SAC components, including Mad2, to unattached kinetochores^{2,3,14,15}. Cells that lack Mad2 arrest in metaphase in the presence of Lat B (Fig. 3a), indicating that Lat B does not activate the SAC. Activation of the Cut1 (Separase) is normally triggered by the rapid destruction of an inhibitory subunit, Cut2 (Securin)^{4,5,16,17}. Cut2 ubiquitination is catalysed by the APC, a specialized E3 ubiquitin ligase, which is also responsible for the destruction of Cdc13 (Cyclin B) and inactivation of the mitotic Cdc2–Cyclin B complex^{4,5,18,19}. Surprisingly, we found that after synchronous release of cells from a G2 block, neither the activation nor inactivation of Cdc2 kinase was affected by Lat B treatment (Fig. 3b); the timing or extent of either Cdc13 or Cut2 destruction was also unaffected (Fig. 3c), suggesting that Lat B does not inhibit APC function. We conclude that Lat B delays cohesin cleavage and sister chromatid separation without activation of the SAC or inhibition of the APC.

To identify components of this new mitotic checkpoint we screened a range of mutants for Lat B sensitivity. Using a plate

assay we found that cells lacking Atf1 (refs 20, 21), a homologue of the human Atf2 transcription factor, were significantly more sensitive to 4 μ M Lat B than the wild type (Fig. 4a). Atf1 is phosphorylated and activated by the Sty1 MAP kinase (also known as SPC1), a homologue of the mammalian p38 MAP kinase^{20–23}. We found that cells lacking the Sty1 MAP kinase were also defective in Lat B-mediated mitotic delay but arrested normally in response to thiabendazole (TBZ), a microtubule inhibitor that activates the SAC (data not shown). Furthermore, Sty1 was phosphorylated and activated by addition of Lat B but not TBZ (Fig. 4b). This resulted in the expression of Atf1-dependent genes such as *pyp2* (Fig. 4c). Induction of Sty1 phosphorylation by Lat B was not observed in a *cps8* mutant, and *cps8* Δ *sty1* double mutants were inviable at all temperatures (data not shown). Importantly, we found that synchronized Δ *atf1* cells completely failed to arrest nuclear division in the presence of Lat B (Fig. 4d) and that these cells lost viability exactly as they passed through mitosis (Fig. 4e). When these cells were re-plated on fresh medium they displayed numerous mitotic abnormalities including displaced and fragmented nuclei and mispositioned division septa, whereas control cells were normal (Fig. 4f). A similar phenotype was observed in *cps8* Δ *sty1* double mutant cells when expression of *sty1* from an episomal plasmid was repressed (data not shown). These data indicate that the stress-activated MAP kinase (SAPK) pathway has an integral role in a new mitotic checkpoint that prevents sister chromatid separation, spindle elongation and commitment to cytokinesis in response to damage to the actin cytoskeleton.

Why should actin depolymerization activate a mitotic checkpoint? The answer probably lies in the cortical ring of actin that encircles the early mitotic nucleus in fission yeast. Astral microtubules form on the spindle pole bodies subsequent to ring formation. The function of astral microtubules in spindle elongation is controversial but it seems likely that these microtubules are essential

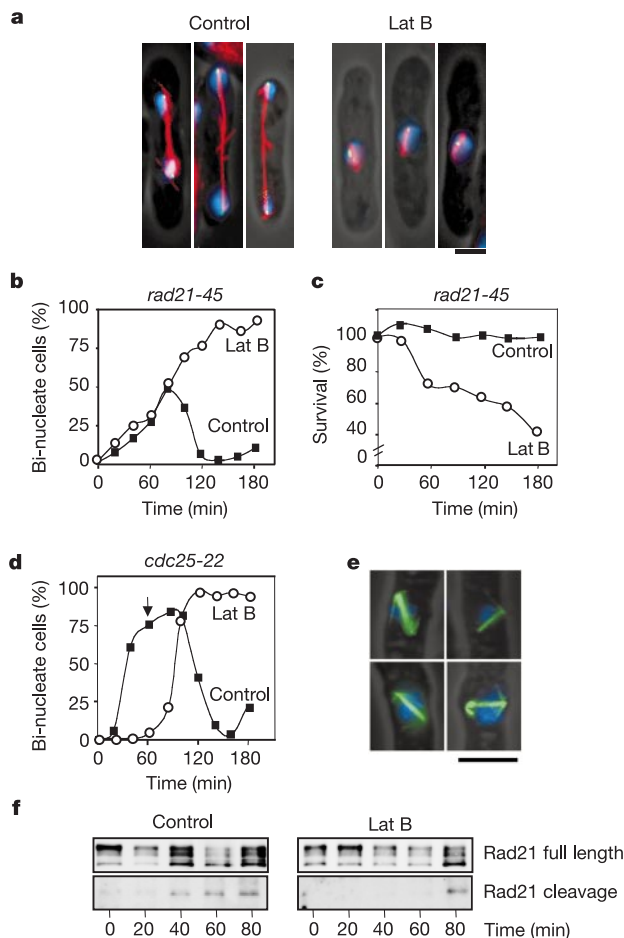


Figure 2 Mitotic actin checkpoint delays the separation of sister chromatids. **a**, Synchronized cells expressing a GFP tag 30 kilobases from centromere 1 (*cen1*–GFP) in the presence (Lat B) or absence (Control) of 10 μ M Lat B, probed with anti-tubulin (red) and stained with DAPI (blue) at 100 min. A single fluorescent dot (yellow) in Lat B-delayed cells represents unseparated centromeres. Scale bar, 4 μ m. **b**, Nuclear division in synchronized *rad21-45* cells with (circles) or without (squares) 10 μ M Lat B. **c**, Percentage survival of synchronized *rad21-45* cells with (circles) or without (squares) 10 μ M Lat B. At the times indicated, cells were extensively washed and viability assessed after 3 days growth on YES plates. **d**, *cdc25-22* cells released from G2/M arrest, in the presence (circles) or absence (squares) of 20 μ M Lat B. **e**, Misoriented spindles of Lat B-treated cells at 60 min (arrow in **d**). Scale bar, 4 μ m. **f**, Cell extracts from **d** probed with anti-Rad21 antibodies. Full-length Rad21 migrates at a relative molecular mass of 105,000–115,000 (M_r 105–115K); the product migrates at about 80K.

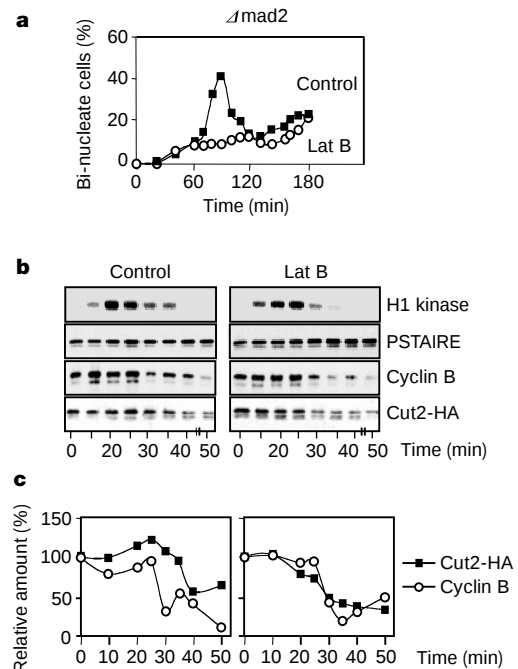


Figure 3 The mitotic actin checkpoint is independent of the SAC and the APC. **a**, Nuclear division in synchronized Δ *mad2* cells with (circles) or without (squares) 10 μ M Lat B. **b**, *cdc25-22 cut2-364::cut2-3HA-LEU2* cells released from G2/M arrest were assayed for Cdc2-associated histone H1 activity and probed for PSTAIRE (Cdc2)³⁰, Cyclin B (Cdc13)¹⁹ or Cut2-HA. **c**, Quantification of Cyclin B (circles) and Cut2-HA (squares) compared to the PSTAIRE control shows that the steady-state level of both decreases by about 50% either with or without Lat B.

to orient the spindle with respect to the long axis of the cell. We assume that the cortical actin ring positions a receptor that stabilizes astral microtubules and that this is important for establishing spindle orientation; disruption of this process results in destabilization of the astral microtubule and spindle misorientation. Consequently, sister chromatid separation, spindle elongation and commitment to cytokinesis are delayed. A possible target of this new checkpoint could be the POLO-related kinase Plo1, which localizes to spindle pole bodies in mitosis and is required both for establishment of the division plane and for mitotic progression^{24,25}. In principle the SAPK pathway could be required for induction of a Plo1 inhibitor, for proper maintenance of cohesion between sister chromatids or for normal interaction of astral microtubules with the cell cortex. Future experiments are needed to distinguish between these possibilities.

Our results uncover the existence of a new mitotic checkpoint, distinct from the SAC, which ensures mitotic spindles are properly oriented before anaphase is allowed to take place. As establishment of asymmetric cell fates during early metazoan development can be dependent on the orientation of spindle position during mitosis^{26,27}, a spindle-orientation checkpoint such as that described here may function to maintain a metaphase arrest until orientation has been established. We predict that bypass of this checkpoint would not only disrupt cell-fate determination, but also contribute to the

uncontrolled proliferation and genome instability of human tumour cells. □

Methods

Cell synchronization

Exponentially growing cells in YEPD medium were synchronized using a 15–40% lactose density gradient at 4 °C, and re-suspended into fresh medium at 30 °C with or without Lat B (10 μ M) (Calbiochem). Exponentially growing cultures of *cdc25-22* cells at 25 °C were arrested in G2 by incubation at 36 °C for 4 h, and then released into mitosis by rapid cooling to 25 °C in the presence of 20 μ M Lat B.

Immunofluorescence and microscopy

Immunofluorescence microscopy procedures were as previously described⁸. The anti-tubulin antibody TAT1 was used after fixation of cells with paraformaldehyde (3.7%) and glutaraldehyde (0.2%) for 1 h. Cen1–GFP expressing cells were fixed with formaldehyde, washed once in phosphate-buffered saline and dried onto coverslips. GFP fusion proteins were observed by using a Zeiss Axiophot microscope using a $\times 64$ 1.4 NA objective. We captured images using Openlab software (Improvision).

Biochemistry

We measured histone H1 kinase activity in *S. pombe* extracts as previously described¹⁸. We performed western blots as previously described²². Anti-Rad21 polyclonal antibody¹², anti-cyclin B polyclonal antibody¹⁹, anti-PSTAIRE monoclonal antibody²⁸, anti-HA antibody (12CA5) (Babco) and anti-Myc antibody (9E10) (Babco) were used at 1:1,000 dilution. Sty1 phosphorylation was measured in a strain containing a 6His-HA-tagged version of Sty1 using an anti-p38 antibody (New England Biolabs) that recognizes the doubly phosphorylated form of Sty1, as previously described²². Autoradiographs were scanned using a molecular dynamics personal densitometer (Amersham).

Received 11 April; accepted 4 May 2001.

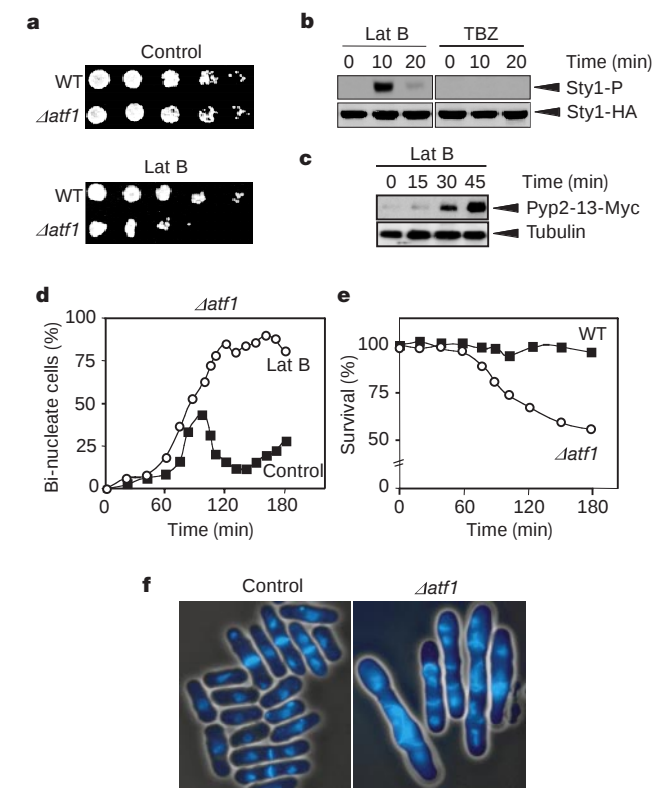


Figure 4 The mitotic actin checkpoint is dependent on the stress-activated Sty1 MAP kinase pathway. **a**, Fivefold serial dilutions of wild-type and $\Delta atf1$ cells spotted onto YES agar with or without 4 μ M Lat B. **b**, *sty1(6His-HA)::ura4* cells were treated with either 10 μ M Lat B or 100 μ M TBZ, and cell extracts probed for phosphorylated Sty1 (Sty1-P) and loading control (Sty1-HA). **c**, *pyp2-13-myc* cells were treated as in **b**, and the extracts probed for Pyp2-13-Myc and for tubulin as loading control. **d**, Nuclear division in synchronized $\Delta atf1$ cells with (circles) or without (squares) 10 μ M Lat B. **e**, Percentage survival of synchronized $\Delta atf1$ cells with (circles) or without (squares) 10 μ M Lat B. At the times indicated, cells were extensively washed and viability assessed after 3 days growth on YES plates. **f**, Synchronized wild-type and $\Delta atf1$ cells incubated in 10 μ M Lat B for 3 h, re-inoculated in fresh medium without Lat B, and grown for another 3 h. Cells were stained with DAPI and calcofluor to reveal the chromatin and division septa, respectively.

- Nicklas, R. B. How cells get the right chromosomes. *Science* **275**, 632–637 (1997).
- Amon, A. The spindle checkpoint. *Curr. Opin. Genet. Dev.* **9**, 69–75 (1999).
- Hardwick, K. G. *et al.* Lesions in many different spindle components activate the spindle checkpoint in the budding yeast *Saccharomyces cerevisiae*. *Genetics* **152**, 509–518 (1999).
- Zachariae, W. & Nasmyth, K. Whose end is destruction: cell division and the anaphase-promoting complex. *Genes Dev.* **13**, 2039–2058 (1999).
- Yanagida, M. *et al.* Control of metaphase-anaphase progression by proteolysis: cyclosome function regulated by the protein kinase A pathway, ubiquitination and localization. *Phil. Trans. R. Soc. Lond. B* **354**, 1559–1569 (1999).
- Marks, J. M. & Hyams, J. S. Localisation of F actin through the cell division cycle of *Schizosaccharomyces pombe*. *Eur. J. Cell Biol.* **39**, 27–32 (1985).
- Spector, I., Shochet, N. R., Kashman, Y. & Groweiss, A. Latrunculin: novel marine toxins that disrupt microfilament organization in cultured cells. *Science* **219**, 493–495 (1983).
- Hagan, I. M. & Hyams, J. S. Forces acting on the fission yeast anaphase spindle. *Cell Motil. Cytoskeleton* **34**, 69–75 (1996).
- Mallavarapu, A., Sawin, K. & Mitchison, T. A switch in microtubule dynamics at the onset of anaphase B in the mitotic spindle of *Schizosaccharomyces pombe*. *Curr. Biol.* **9**, 1423–1426 (1999).
- Nabeshima, K. *et al.* Dynamics of centromeres during metaphase-anaphase transition in fission yeast: Dis1 is implicated in force balance in metaphase bipolar spindle. *Mol. Cell* **9**, 3211–3225 (1998).
- Ishiguro, J. & Kobayashi, W. An actin point-mutation neighboring the 'hydrophobic plug' causes defects in the maintenance of cell polarity and septum organization in the fission yeast *Schizosaccharomyces pombe*. *FEBS Lett.* **392**, 237–241 (1996).
- Birkenbihl, R. P. & Subramani, S. The rad21 gene product of *Schizosaccharomyces pombe* is a nuclear, cell cycle-regulated phosphoprotein. *J. Biol. Chem.* **270**, 7703–7711 (1995).
- Tomonaga, T. *et al.* Characterization of fission yeast cohesin: essential anaphase proteolysis of Rad21 phosphorylated in the S phase. *Genes Dev.* **14**, 2757–2770 (2000).
- Chen, R. H., Waters, J. C., Salmon, E. D. & Murray, A. W. Association of spindle assembly checkpoint component XMad2 with unattached kinetochores. *Science* **274**, 242–246 (1996).
- He, X., Patterson, T. E. & Sazer, S. The *Schizosaccharomyces pombe* spindle checkpoint protein Mad2p blocks anaphase and genetically interacts with the anaphase-promoting complex. *Proc. Natl Acad. Sci. USA* **94**, 7965–7970 (1997).
- Funabiki, H. *et al.* Cut2 proteolysis required for sister-chromatid separation in fission yeast. *Nature* **381**, 438–441 (1996).
- Uhlmann, F. *et al.* Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. *Cell* **103**, 375–386 (2000).
- Moreno, S., Hayles, J. & Nurse, P. Regulation of p34cdc2 protein kinase during mitosis. *Cell* **58**, 361–372 (1989).
- Alfà, C. E., Ducommun, B., Beach, D. & Hyams, J. S. Distinct nuclear and spindle pole body population of cyclin-cdc2 in fission yeast. *Nature* **347**, 680–682 (1990).
- Shiozaki, K. & Russell, P. Conjugation, meiosis, and the osmotic stress response are regulated by Spc1 kinase through Atf1 transcription factor in fission yeast. *Genes Dev.* **10**, 2276–2288 (1996).
- Wilkinson, M. G. *et al.* The Atf1 transcription factor is a target for the Sty1 stress activated MAP kinase pathway in fission yeast. *Genes Dev.* **10**, 2289–2301 (1996).
- Millar, J. B. A., Buck, V. & Wilkinson, M. G. Pyp1 and Pyp2 PTases dephosphorylate an osmosensing MAP kinase controlling cell size at division in fission yeast. *Genes Dev.* **9**, 2117–2130 (1995).
- Shiozaki, K. & Russell, P. Cell cycle control linked to extracellular environment by MAP kinase pathway in fission yeast. *Nature* **378**, 739–743 (1995).
- Ohkura, H., Hagan, I. M. & Glover, D. M. The conserved *Schizosaccharomyces pombe* kinase plo1, required to form a bipolar spindle, the actin ring, and septum, can drive septum formation in G1 and G2 cells. *Genes Dev.* **9**, 1059–1073 (1995).
- Tanaka, K. *et al.* The role of Plo1 kinase in mitotic commitment and septation in *Schizosaccharomyces pombe*. *EMBO J.* **20**, 1259–1270 (2001).

26. Knoblich, J. A. Mechanisms of asymmetric cell division during animal development. *Curr. Opin. Cell Biol.* **9**, 833–841 (1997).
27. Lu, B., Jan, L. Y. & Jan, Y. N. Asymmetric cell division: lessons from flies and worms. *Curr. Opin. Genet. Dev.* **8**, 392–399 (1998).
28. Yamashita, M., Yoshikuni, M., Hirai, T., Fukada, S. & Nagahama, Y. A monoclonal antibody against the PSTAIR sequence of p34^{cdc2} catalytic subunit of maturation-promoting factor and key regulator of the cell cycle. *Dev. Growth Differ.* **33**, 617–624 (1991).

Acknowledgements

We thank V. Amin and K. Adams for technical assistance; F. Uhlmann and K. Nasmyth for communicating results before publication; and T. Matsumoto, S. Subramani and M. Yanagida for strains and reagents. We also thank V. Buck, L. Johnston, S. Jensen and D. Mulhivill for comments and critical reading of the manuscript, and members of the Hyams and Millar laboratories for their encouragement and support. This work was supported by a Wellcome Trust grant to J.S.H. and by a MRC grant to J.B.A.M.

Correspondence and requests for materials should be addressed to J.B.A.M. (e-mail: jmillar@nimr.mrc.ac.uk).

Cdc6 cooperates with Sic1 and Hct1 to inactivate mitotic cyclin-dependent kinases

Arturo Calzada*, María Sacristán*, Elisa Sánchez & Avelino Bueno

Instituto de Microbiología-Bioquímica e Instituto de Biología Molecular y Celular del Cáncer, Departamento de Microbiología y Genética, Campus Miguel de Unamuno, Universidad de Salamanca, CSIC 37007 Salamanca, Spain

* These authors contributed equally to this work

Exit from mitosis requires the inactivation of mitotic cyclin-dependent kinases (CDKs). In the budding yeast, *Saccharomyces cerevisiae*, inactivation of CDKs during late mitosis involves degradation of B-type cyclins as well as direct inhibition of cyclin-CDK complexes by the CDK-inhibitor protein Sic1 (refs 1–3). Several striking similarities exist between Sic1 and Cdc6, a DNA replication factor essential for the formation of pre-replicative complexes at origins of DNA replication^{4–9}. Transcription of both genes is activated during late mitosis by a process dependent on Swi5 (ref. 10). Like Sic1, Cdc6 binds CDK complexes *in vivo*^{11,12} and downregulates them *in vitro*¹¹. Here we show that Cdc6, like Sic1, also contributes to inactivation of CDKs during late mitosis in *S. cerevisiae*. Deletion of the CDK-interacting domain of Cdc6 does not inhibit the function of origins of DNA replication during S phase, but instead causes a delay in mitotic exit; this delay is accentuated in the absence of Sic1 or of cyclin degradation. By contributing to mitotic exit and inactivation of CDKs, Cdc6 helps to create the conditions that are required for its subsequent role in the formation of pre-replicative complexes at origins of DNA replication.

We first examined whether Cdc6 acts as a CDK inhibitor when overexpressed *in vivo*. We constructed *S. cerevisiae* strains in which expression of *CDC6*, or a mutated version lacking the amino-terminal 47 amino acids ($\Delta 47cdc6$)¹³, were regulated by the galactose-inducible *GALI-10* promoter. Exponential cultures were grown in the absence of galactose, and cells were then arrested in G2/M with the microtubule-depolymerizing drug nocodazole. Expression of *CDC6* or $\Delta 47cdc6$ was then induced by adding galactose to the medium. Samples were taken at regular intervals and processed for CDK kinase assays. Overexpression of wild-type Cdc6 inhibited the activity of the cyclin B-CDK complex Clb2-Cdc28, whereas the induction of $\Delta 47cdc6$ did not affect Clb2-Cdc28 activity, as expected (Fig. 1a). Cdc6-mediated inactivation was independent of Sic1-mediated Clb2-Cdc28 inhibition

(Fig. 1b). Similarly, inhibition of Cdc28 activity by Cdc6 did not require Hct1 (ref. 2), an activator of cyclin B degradation by means of the anaphase promoting complex (APC; Fig. 1c). Together with the *in vitro* ability of Cdc6 to inhibit Cdc28 (ref. 11), these results indicate that Cdc6 can function as a CDK inhibitor, and that this function is exerted through the Cdc28-interacting domain present in the first 47 amino acids of the Cdc6 protein.

Null mutants of *HCT1* ($\Delta hct1$) are viable and their exit from mitosis is only moderately delayed compared with wild-type cells². This is probably because defects in cyclin proteolysis are compensated by direct CDK inhibition². If Cdc6, together with Sic1, inhibits CDKs as cells exit mitosis it would be expected that *cdc6* mutant cells would show defects in exiting this stage of the cell cycle. We first confirmed that the $\Delta 47cdc6$ strain did not exhibit any defects in DNA replication. Rates of plasmid loss are very similar between $\Delta 47cdc6$ and $\Delta 47cdc6 \Delta hct1$ strains and appropriate wild-type or $\Delta hct1$ controls (per cent of loss rates, mean $\pm$ s.d.: wild type, 1.81 ± 0.16 ; $\Delta 47cdc6$, 1.51 ± 0.23 ; $\Delta hct1$, 2.72 ± 0.31 ; $\Delta 47cdc6 \Delta hct1$, 2.28 ± 0.18). Second, we used two-dimensional gel analyses to show that $\Delta 47cdc6$ cells fire replication origins as efficiently as wild-type cells (as shown for *ARS1* in Supplementary Information Fig. 1).

To study whether $\Delta 47cdc6$ mutant cells were defective in their exit from mitosis, we first used nocodazole block-and-release experiments, because this allowed us to analyse fully all the events from mitosis until entry into G1 phase. In keeping with defects in the exit from mitosis, the decline of cells with mitotic spindles was delayed in $\Delta 47cdc6$ and $\Delta hct1$ cells compared with *CDC6* wild-type cells. Also, Clb2-associated kinase activity and Clb2 protein levels declined slightly more rapidly in the wild-type control than in mutant cells (see Supplementary Information Fig. 2). We then analysed exit from mitosis in cells synchronized in telophase, by arresting a *cdc15* mutant at the restrictive temperature of 37 °C (ref. 14). Thus, exponential cultures of *cdc15-2* and isogenic *cdc15-2 \Delta 47cdc6* and *cdc15-2 \Delta hct1* strains were blocked at 37 °C for 4 h. Cells were then released from the block and further incubated at 20 °C; samples were taken at regular intervals and processed for immunofluorescence, analysed by western blot and assayed for kinase activity (Fig. 2). In both *cdc15-2 \Delta 47cdc6* and *cdc15-2 \Delta hct1* mutants, the rate of disappearance of elongated spindles was delayed compared with the control (Fig. 2b). Consistent with this delay, on release from the telophase arrest, activity of Clb2-associated H1 kinase declined faster in the *cdc15-2* control than in *cdc15-2 \Delta 47cdc6* or *cdc15-2 \Delta hct1* (Fig. 2b). Levels of Clb2 protein declined both in the *cdc15-2* control and in *cdc15-2 \Delta 47cdc6* mutant cells. In keeping with previous observations², Clb2 was stable in the strain with *HCT1* deleted.

We also found that the $\Delta 47cdc6$ mutant is synthetically lethal when combined with mutations in the *CDC23* gene, which encodes a component of the APC^{15–17}, just as observed previously for $\Delta sic1$ (ref. 2). Several $\Delta 47cdc6 cdc23-1$ haploid strains were selected at 25 °C and found to be inviable above 28 °C, a temperature at which $\Delta 47cdc6$ and *cdc23-1* single mutants grow at a similar rate as the wild type (see Supplementary Information Fig. 3). Synthetic lethality of $\Delta 47cdc6$ with an APC mutant suggests either that the APC is involved in the regulation of Cdc6 protein stability, or that Cdc6 cooperates with the APC in promoting progression through mitosis, in which APC also participates^{2,15–17}. As proteolysis of Cdc6 is not regulated by the APC¹⁸, our data, together with the synthetic lethality described for *cdc6-1 cdc14-1* mutants¹⁹, suggest a scheme in which Cdc6 cooperates with Sic1 in CDK inhibition. We also found that $\Delta 47cdc6$ is synthetically lethal at intermediate temperatures when combined with other mutants defective in mitotic exit, such as *cdc14-1*. Several haploid $\Delta 47cdc6 cdc14-1$ strains were selected and found to be inviable at 31 °C (see Supplementary Information Fig. 3). This synthetic lethality was comparable to the genetic interactions observed in *cdc14-1 \Delta sic1* and *cdc14-1 \Delta hct1* double mutants^{19,20}. $\Delta 47cdc6 cdc14-1$ double mutants arrested at 31 °C as

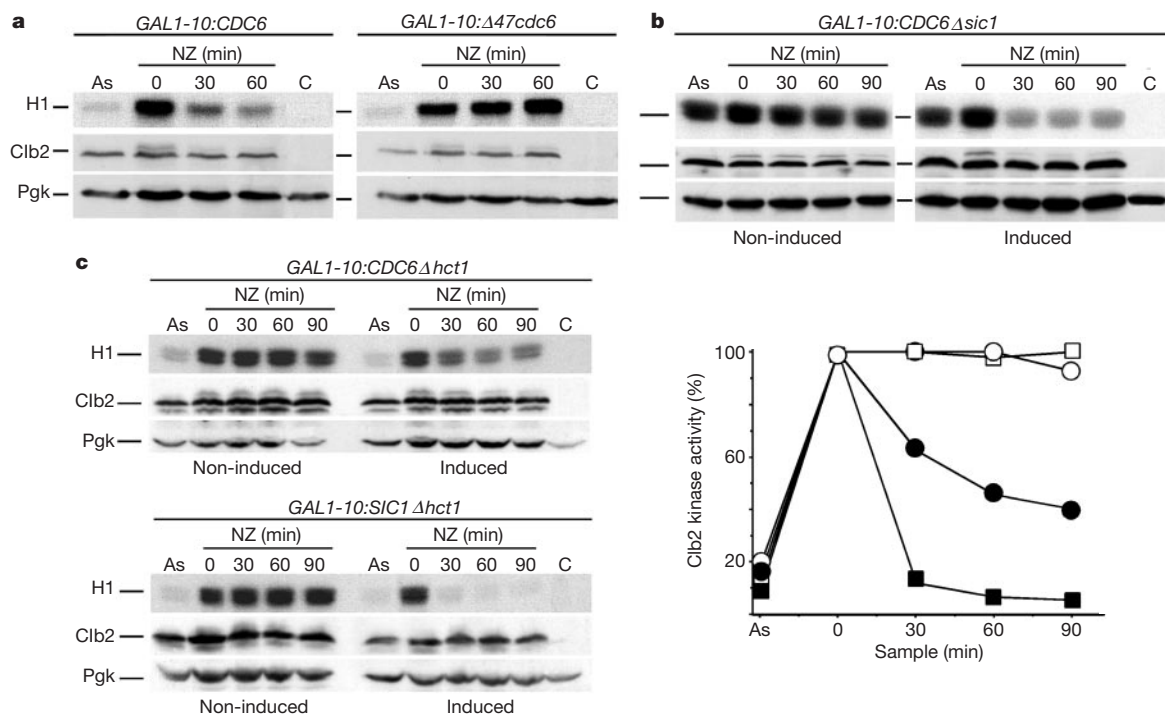


Figure 1 Cdc6 promotes Clb2-associated kinase inhibition by a mechanism independent of Clb2 degradation. Cultures of congenic wild type (**a**), $\Delta sic1$ (**b**) and $\Delta hct1$ (**c**) cells expressing *CDC6*, $\Delta 47cdc6$ or *SIC1* under the control of the galactose-inducible promoter *GAL1-10* were arrested in mitosis with nocodazole (NZ) before the addition of galactose. Samples were taken at regular intervals and processed for Clb2-associated H1

kinase activity assays and western analysis (Pgk as a loading control). Asynchronous (As) and untagged Clb2 (C) controls are shown for reference. In **c**, circles indicate *GAL1-10:CDC6*, squares *GAL1-10:SIC1*, open symbols non-induced and closed symbols induced samples.

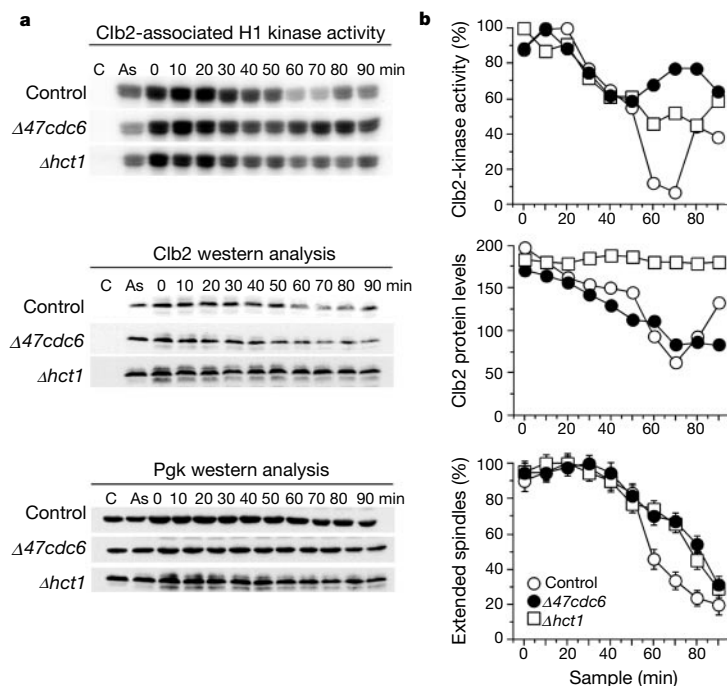


Figure 2 Exit from mitosis is delayed in $\Delta 47cdc6$ mutant cells compared with *CDC6* wild-type cells. Cultures of *cdc15-2* (control), *cdc15-2 Δ47cdc6* ($\Delta 47cdc6$) and *cdc15-2 Δhct1* ($\Delta hct1$) strains were synchronized in telophase by incubating them at 37 °C for 4 h. Cultures were then released at 20 °C and samples taken at indicated intervals and

processed for Clb2-associated H1 kinase activity assays (**a**, top panels), Clb2 and Pgk western analysis (**a**, middle and lower panels) and tubulin staining (**b**, lower plot) to measure the exit from mitosis. Plots of the relative H1 kinase activity and Clb2 protein abundance (normalized to Pgk levels) are shown in **b**.

single, elongated budded cells with two distinct nuclei (not shown), similar to *cdc14-1* mutant cells at 37 °C (ref. 21) and consistent with a block late in mitosis. These genetic interactions suggest that the Cdc28-interaction domain at *CDC6* participates in mitotic exit.

If Cdc6 cooperates with Hct1 and Sic1 to promote mitotic exit, the role of Cdc6 in this process might be easier to observe in cells lacking Hct1 or Sic1. We also characterized, therefore, the phenotype of the double mutant $\Delta 47cdc6 \Delta hct1$. These mutant cells, although viable, were larger (Fig. 3b) and grew slower than wild-type or single-mutant controls, showing a significant lag in late mitosis (Fig. 3). Consistent with this lag, asynchronously growing cells showed a marked increase in 2C DNA cell contents as compared with wild-type cells (Fig. 3a) and had approximately 50% of the levels of Clb2–Cdc28 kinase activity to nocodazole-arrested cells (Fig. 3d). Furthermore, tubulin staining revealed that 80.1 $\pm$ 2.71% (mean $\pm$ s.d.) of these cells had an extended mitotic spindle reminiscent of defects late in mitosis (Fig. 3c), a percentage significantly higher than those observed in wild-type (14.1 $\pm$ 0.76%), $\Delta 47cdc6$ (16.3 $\pm$ 0.97%) and $\Delta hct1$ (26.5 $\pm$ 1.32%) controls (this defect is rescued by *SIC1*; see Supplementary Information Fig. 4). A telophase block-and-release experiment was then performed to compare the Clb2-associated kinase activity, and also to measure the extended spindle index in *cdc15-2 \Delta hct1* and *cdc15-2 \Delta 47cdc6 \Delta hct1* strains (Fig. 3e). After a 4-h block at 37 °C, cells were released and further incubated at 25 °C, taking samples at regular intervals. Consistent with a role for Cdc6 in the exit from mitosis, in

cdc15-2 \Delta 47cdc6 \Delta hct1 cells both the rate of disappearance of telophase spindles and the drop in Clb2–Cdc28 kinase activity was significantly delayed compared with the *cdc15-2 \Delta hct1* control (Fig. 3e).

In a similar manner, we examined the effects of combining $\Delta 47cdc6$ with deletion of the *SIC1* gene. We found that the double mutant $\Delta 47cdc6 \Delta sic1$ is inviable at all temperatures. To study the double-mutant phenotype, a conditional $\Delta 47cdc6 \Delta sic1$ *GAL1-10:cdc6-1* haploid strain was constructed. This strain grew in galactose-based medium at 25 °C. Nevertheless, repression of the *GAL1-10* promoter at 37 °C induced the appearance of large budded cells with replicated DNA, two nuclei and extended mitotic spindles (Fig. 4), consistent with a defect in late anaphase. This result provides further evidence supporting a role for Cdc6 in the exit from mitosis, in combination with Sic1.

CDC6 is essential in the formation of pre-replicative complexes at origins of DNA replication^{5,22}. It is known that the periodic transcription of the *CDC6* gene in rapidly proliferating cells starts late in mitosis^{23,24} and this is correlated with the appearance of the Cdc6 initiation protein⁷. We have discovered a major reason explaining why Cdc6 is synthesized at this stage of the cell cycle: namely, that it contributes to the inactivation of CDK at the end of mitosis, cooperating with Sic1 in the direct inhibition of cyclin B–CDK complexes (Fig. 4c).

Our findings show that in budding yeast Cdc6 has a dual role, in both chromosome replication and in mitosis, and it will be of

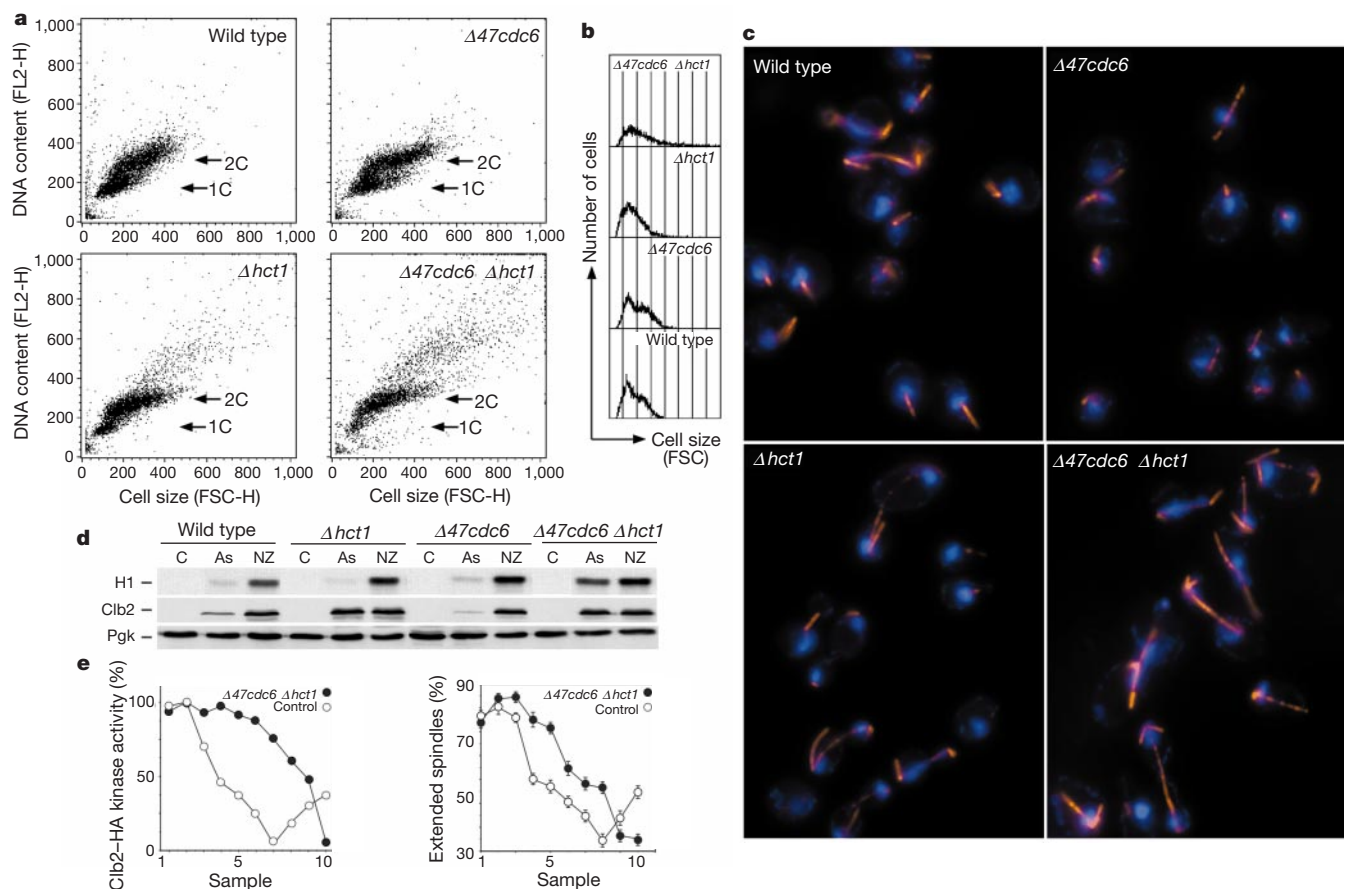


Figure 3 The 47 amino-terminal truncation in *CDC6* confers a strong mitotic delay to $\Delta hct1$ cells. Dot plots (**a**) of DNA content versus cell size and histograms (**b**) of cell size from samples of asynchronous cultures (28 °C) of wild-type, $\Delta 47cdc6$, $\Delta hct1$ and $\Delta hct1 \Delta 47cdc6$ isogenic strains. **c**, Micrographs of these cells co-stained with anti-tubulin and DAPI. **d**, Immunoprecipitated Clb2-associated H1 kinase activity of the same cultures (As) as compared to nocodazole-arrested cells (NZ) and an untagged Clb2 control (C). Clb2 and Pgk western blots are shown. **e**, A telophase block-and-release experiment of *cdc15-2 \Delta hct1* (control) and *cdc15-2 \Delta hct1 \Delta 47cdc6* strains. Plots of Clb2–HA kinase activity and extended spindles are shown. Samples are telophase-arrested cells (1), or taken every 10 min after the release (2–10).

as compared to nocodazole-arrested cells (NZ) and an untagged Clb2 control (C). Clb2 and Pgk western blots are shown. **e**, A telophase block-and-release experiment of *cdc15-2 \Delta hct1* (control) and *cdc15-2 \Delta hct1 \Delta 47cdc6* strains. Plots of Clb2–HA kinase activity and extended spindles are shown. Samples are telophase-arrested cells (1), or taken every 10 min after the release (2–10).

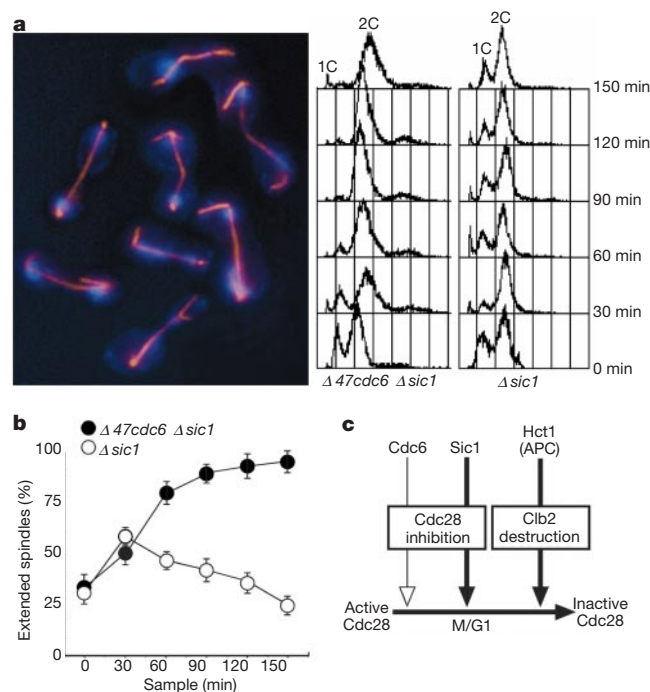


Figure 4 $\Delta sic1 \Delta 47cdc6$ double-mutant cells are defective in the exit from mitosis. Asynchronously growing cultures of $\Delta 47cdc6 \Delta sic1$ GAL1-10:*cdc6-1* and $\Delta sic1$ GAL1-10:*cdc6-1* strains were repressed and shifted to 37 °C to analyse DNA content (a) and extended spindle morphology (b) during a time-course experiment. $\Delta 47cdc6 \Delta sic1$ cells from the last time point (a) were co-stained with anti-tubulin antibodies and DAPI to show spindle and nuclear morphology, respectively. c, A model for the relative importance of Cdc6, Sic1 and Hct1 in CDK inactivation at the exit from mitosis.

considerable interest to see whether the same is true for homologues of Cdc6 in other species of eukaryote. □

Methods

Strains and cell cycle control

All strains used in this work were backcrossed three times with a 15Dau strain¹². All of them carried a *CLB2HA* allele¹². They were constructed by tetrad analysis, checking appropriate segregation of markers after each cross genetically, by Southern blot and/or polymerase chain reaction (PCR) and segregation of *CLB2HA* by western blot. To construct the $\Delta 47cdc6$ strain that carries this allele as the only copy of the *CDC6* allele, two PCRs were performed. We generated a fragment with 596 base pairs (bp) (flanked by *EcoRI*–*NdeI* sites) containing the 5′-flanking sequence upstream from the *CDC6* ATG (oligonucleotides 5′-TAGAATTCGAGGCTTCTCGAGG-3′ and 3′-CATGGTATATGC AAGGTATACAT-5′) and a 633-bp fragment (flanked by *NdeI* and *XbaI* sites) from the 145–778 bp of the *CDC6* open reading frame (oligonucleotides 5′-TACATATGTTTGGC TCACAGTCT-3′ and 3′-CCATACCGATTATCAGATCTATAC-5′). The two fragments were ligated using the *NdeI* sites and cloned into a pGEM-T vector. The internal *MluI*–*XbaI* fragment was fully sequenced and used to replace the wild-type *MluI*–*XbaI* fragment contained in the genomic *BamHI*–*HindIII* 5.5-kilobase (kb) fragment. A *URA3* gene was introduced as marker into the *CDC6* downstream from the *EcoRI* site and the resulting *BamHI*–*HindIII* fragment was used to transform 15Dau cells. The resulting genomic constructions were tested by PCR and sequencing.

Yeast strains were grown in rich YEP medium (1% yeast extract, 2% peptone) containing 2% glucose. For block-and-release experiments, cells were grown in YEP with 2% glucose at 28 °C and synchronized by adding 15 $\mu\text{g ml}^{-1}$ nocodazole for 150 min, except where indicated. Cells were collected by centrifugation and released in nocodazole-free medium for 90 min at 28 °C. Overexpression experiments were done with cells growing in YEP medium with 2% raffinose at 32 °C; after 120 min of nocodazole blocking, 2.5% galactose was added to induce (or 2% glucose to repress) expression from the *GAL1* promoter. The cell cycle position was monitored by standard FACS analysis. Analysis of plasmid loss rate was performed as described²⁵.

Kinase analysis and western blot

Cell extracts were prepared as described¹² and protein concentrations were determined using BCA Protein Assay Kit (Pierce). Immunoprecipitation and immunoblotting were carried out with anti-haemagglutinin (HA) (12CA5, Roche) and anti-Pgk (mouse monoclonal, Molecular Probes) antibodies. Kinase assays were performed with H1 histone

(Calbiochem) as substrate on immunoprecipitates at 25 °C as described¹², and quantified with a phosphorimager. H1 histone signals were normalized to the nocodazole-arrest level.

Microscopy

To evaluate exit from mitosis, elongated spindles were counted and plotted. Staining of microtubules was performed as described²⁶ using anti-tubulin antibody TAT-1, and DNA was visualized after staining with 4,6-diamidino-2-phenylindole (DAPI). Fluorescence images were collected using a Leica Q550CW microscope with a $\times 63$ objective and a Sensys CCD (charge-coupled device) camera (Photometrics). More than 300 cells were examined for each time point, and the experiment was repeated three times, to gain an estimate of error.

Received 7 February; accepted 18 June 2001.

- Donovan, J. D., Toyn, J. H., Johnson, A. L. & Johnston, L. H. p40SDB25, a putative CDK inhibitor, has a role in the M/G1 transition in *Saccharomyces cerevisiae*. *Genes Dev.* **8**, 1640–1653 (1994).
- Schwab, M., Schulze-Lutum, A. & Seufert, W. Yeast Hct1 is a regulator of Clb2 cyclin proteolysis. *Cell* **90**, 683–693 (1997).
- Visintin, R., Prinz, S. & Amon, A. CDC20 and CDH1: a family of substrate-specific activators of APC-dependent proteolysis. *Science* **17**, 460–463 (1997).
- Liang, C., Weinreich, M. & Stillman, B. ORC and Cdc6p interact and determine the frequency of initiation of DNA replication in the genome. *Cell* **81**, 667–676 (1995).
- Cocker, J. H., Piatti, S., Santocanale, C., Nasmyth, K. & Diffley, J. F. X. An essential role for the Cdc6p protein in budding yeast pre-replicative complexes. *Nature* **379**, 180–182 (1996).
- Aparicio, O. M., Weinstein, D. M. & Bell, S. P. Components and dynamics of DNA replication complexes in *S. cerevisiae*: redistribution of MCM proteins and Cdc45p during S phase. *Cell* **91**, 59–69 (1997).
- Detweiler, C. S. & Li, J. Cdc6p establishes and maintains a state of replication competence during G1 phase. *J. Cell Sci.* **110**, 753–763 (1997).
- Donovan, S., Harwood, J., Drury, L. S. & Diffley, J. F. Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. *Proc. Natl Acad. Sci. USA* **94**, 5611–5616 (1997).
- Tanaka, T., Knapp, D. & Nasmyth, K. Loading of an Mcm protein onto DNA replication origins is regulated by Cdc6p and CDKs. *Cell* **90**, 649–660 (1997).
- Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297 (1998).
- Elasser, S., Lou, F., Wang, B., Campbell, J. L. & Jong, A. Interaction between yeast Cdc6 protein and B-type cyclin/Cdc28 kinases. *Mol. Biol. Cell* **7**, 1723–1735 (1996).
- Calzada, A., Sánchez, M., Sánchez, E. & Bueno, A. The stability of the Cdc6 protein is regulated by CDK/cyclin B complexes in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **275**, 9734–9741 (2000).
- Drury, L. S., Perkins, G. & Diffley, J. F. X. The Cdc4/34/53 pathway targets Cdc6p for proteolysis in budding yeast. *EMBO J.* **16**, 5966–5976 (1997).
- Yeong, F. M., Lim, H. H., Padmashree, C. G. & Surana, U. Exit from mitosis in budding yeast: biphasic inactivation of the Cdc28-Clb2 mitotic kinase and the role of Cdc20. *Mol. Cell* **5**, 501–511 (2000).
- Irniger, S., Piatti, S., Michaelis, C. & Nasmyth, K. Genes involved in sister chromatid separation are needed for B-type cyclin proteolysis in budding yeast. *Cell* **81**, 269–278 (1995).
- King, R. W. *et al.* 20S complex containing CDC27 and CDC16 catalyzes the mitosis-specific conjugation of ubiquitin to cyclin B. *Cell* **81**, 279–288 (1995).
- Peters, J. M. Subunits and substrates of the anaphase-promoting complex. *Exp. Cell Res.* **248**, 339–349 (1999).
- Pichler, S., Piatti, S. & Nasmyth, K. Is the yeast Anaphase Promoting Complex needed to prevent re-replication during G2 and M phases? *EMBO J.* **16**, 5988–5997 (1997).
- Yuste-Rojas, M. & Cross, F. R. Mutations in *CDC14* result in high sensitivity to cyclin gene dosage in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **263**, 60–72 (2000).
- Visintin, R. *et al.* The phosphatase Cdc14 triggers mitotic exit by reversal of Cdk-dependent phosphorylation. *Mol. Cell* **2**, 709–718 (1998).
- Culotti, J. & Hartwell, L. H. Genetic control of the cell division cycle in yeast III. Seven genes controlling nuclear division. *Exp. Cell Res.* **67**, 389–401 (1971).
- Mizushima, T., Takahashi, N. & Stillman, B. Cdc6p modulates the structure and DNA binding activity of the origin recognition complex *in vitro*. *Genes Dev.* **14**, 1631–1641 (2000).
- Zwerschke, W., Rottjakob, H.-W. & Kuntzel, H. The *Saccharomyces cerevisiae* *CDC6* gene is transcribed at late mitosis and encodes an ATP/GTPase controlling S phase initiation. *J. Biol. Chem.* **269**, 23352–23356 (1994).
- Piatti, S., Lengauer, C. & Nasmyth, K. Cdc6 is an unstable protein whose de novo synthesis in G1 is important for the onset of S phase and for preventing a 'reductional' anaphase in the budding yeast *Saccharomyces cerevisiae*. *EMBO J.* **14**, 3788–3799 (1995).
- Hogan, E. & Koshland, D. Addition of extra origins of replication to a minichromosome suppressed its mitotic loss in *cdc6* and *cdc14* mutants of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **89**, 3098–3102 (1992).
- Bähler, J. & Pringle, J. R. Pom1p, a fission yeast protein kinase that provides positional information for both polarized growth and cytokinesis. *Genes Dev.* **12**, 1356–1370 (1998).

Supplementary information is available from Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank J. Correa, K. Labib, S. Moreno, L. Pelloquin and the ASP laboratory for many discussions, and K. Labib for critically reading this manuscript. We are grateful to R. Basco, J. Correa, J. Diffley, S. Ufano and C. Vázquez for yeast strains and anti-Pgk antibody. We also thank N. Skinner for editing the text. This research was supported by CICYT and PGC grants to A.B.

Correspondence and requests for materials should be addressed to A.B. (e-mail: abn@gugu.usal.es).

Immuno-this, immuno-that

Recent introductions in the fields of immunology and immunoassay.

BD Biocoat

BD Biosciences www.bdbiosciences.com

Sandwich plates

Available with anti-mouse IgG, anti-rabbit IgG or anti-human IgG-coated antibody, Biocoat secondary antibody assay plates are pre-blocked with protease-free bovine serum albumin and screened for low background and minimal cross-reactivity. Applications include sandwich/competitive immunoassays and measurement of IgG titre levels.

Streptavidin Immobilizer

Flowgen www.flow.co.uk

Ties that bind

Flowgen's new Streptavidin Immobilizer plates and strips are ready to use, requiring no activation or blocking steps, and can detect various types of biotinylated biomolecules, including biotinylated oligonucleotides, peptides, proteins and antibodies. There is no surface leaching and the plates and strips are stable at room temperature for at least 12 months. The molecular weight of streptavidin is approximately 60,000 and its isoelectric point close to 5. Applications include DNA hybridization assays, ELISAs, analysis of DNA binding proteins and PCR-ELISAs.

ImmunEasy

Qiagen www.qiagen.com

Mouse adjuvant

ImmunEasy adjuvant is designed specifically to induce high antibody titres in mice. It comes as a ready-to-use aqueous suspension to be mixed with antigen in one step. Antibody titres are induced by CpG DNA based on a natural reaction, and immunized animals show no symptoms of toxicity. Boosting can be performed after two weeks. ImmunEasy can be used with most types of antigens, including peptides, soluble proteins, virus-like particles, whole-cell preparations and polysaccharide conjugates.

MabClone M86

Alexis Biochemicals www.alexis-corp.com

Made to measure α -Gal expression

Mab Clone M86 is used in the study of human xenotransplantation and related research to measure α -Gal epitope expression on cells, glycolipids, and glycoproteins and to characterise hyperacute rejections (HAR) in organ and tissue transplantation. It can also be employed to conduct cytotoxicity assays for α -Gal specific pathways, monitor transplantation experiments and screen inhibitors of antibody-



It's in the bag: more coated microplates from BD Biosciences.

induced cytotoxicity. An IgM isotype that mimics *in vivo* reaction, Mab M86 is highly specific for synthetic and naturally produced Gal α 1,3Gal epitopes on glycoproteins and glycolipids, and exhibits broad species cross-reactivity. It does not cross-react with β -Gal glycoproteins or BSA.

Human cytochrome c ELISA kit

Bender Medsystems

www.bendermedsystems.com

Apoptosis detector

This kit can be used for quantitative detection of human cytochrome c in cell culture lysates, human whole blood, or serum to detect activation of apoptotic programs in cells. Recent research with serum from patients with haematological malignancies indicates that serum cytochrome c monitoring might also serve as a clinical marker to indicate the onset of apoptosis and cell turnover *in vivo*.

Retrievit

Innogenex www.innogenex.com

Solutions for antigen retrieval problems

Retrievit is a range of target retrieval solutions designed to enhance staining of formalin-fixed tissues. Tissue sections are heated in the appropriate solution for about 15 minutes and cooled for 20 minutes prior to immunostaining. The solutions contain an organic buffer, a strong ion chelator and a mild chaotropic reagent configured to improve *in situ* hybridization signals and expand the range of antibodies that can be used in immunohistochemistry. False negative and weak staining caused by formalin fixation and tissue process-

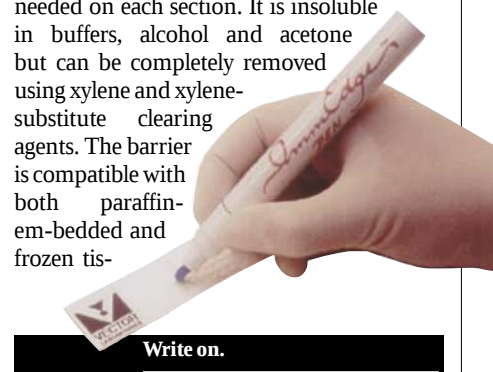
ing can be reversed. The solutions are offered with different pHs suitable for different antigen-antibody pairs. Acidic Retrievit-2 (pH 2) and Retrievit-4 (pH 4) solutions are recommended for retrieval of nuclear antigens, and the neutral-range, Retrievit-6 (pH 6) and Retrievit-8 (pH 8), are suitable for cytoplasmic antigens. Antigens located in the extracellular matrix or on the membrane surface are retrieved most effectively by alkaline Retrievit-8 (pH 8) and Retrievit-10 (pH 10) solutions.

ImmEdge pen

Vector Laboratories www.vectorlabs.com

This one has the edge

An improved version of the ImmEdge hydrophobic barrier pen, this new model contains no 'ozone-depleting' fluorocarbons or chlorocarbons. The pen provides a heat-stable, water-repellent, light-coloured barrier that keeps staining reagents localized on the tissue section, prevents mixing of reagents when differentially staining two sections on the same slide, and minimizes the amount of reagent needed on each section. It is insoluble in buffers, alcohol and acetone but can be completely removed using xylene and xylene-substitute clearing agents. The barrier is compatible with both paraffin-embedded and frozen tis-



Write on.

sue sections mounted on glass slides. The pens are supplied in sets of two.

AP- and FITC-conjugated antibodies

Invitrogen www.invitrogen.com

Monoclonal antibodies for epitope detection

Horseradish peroxidase (HRP), alkaline phosphatase (AP) and fluorescein isothiocyanate (FITC) conjugated monoclonal antibodies are available for western blot or immunofluorescence assays. These antibodies specifically recognize some of the most widely used fusion tags, including polyhistidine (6× His), *c-myc* and V5. Using a conjugated antibody for direct immunodetection eliminates the need for secondary polyclonal antibody, allowing more rapid detection and less background noise.

Human serum IFN-α ELISA

PBL Biomedical Laboratories

www.pblbio.com

Eliminate false positives

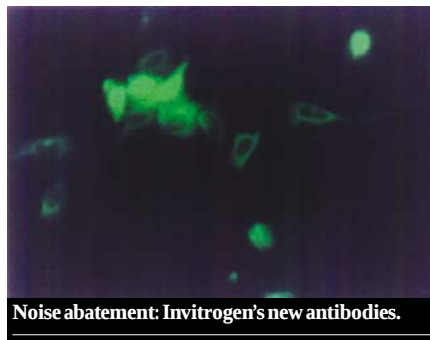
Based on PBL's Human IFN-α Multi-Species ELISA kit, the Human IFN-α Serum Samples ELISA kit has been created specifically for the measurement of human interferon-α in human serum samples without being affected by heterophile antibodies. The kit detects all alpha subtypes except for alpha F and has been specially enhanced to eliminate false positives caused by heterophile antibodies.

β-APP antibody

Zymed Laboratories www.zymed.com

Pointers to axonal injury

β-Amyloid precursor protein exists in the brain as three major isoforms (β-APP695, β-APP751 and β-APP770). Zymed have now introduced C-terminal-specific antibody to β-APP695. Most widely used β-APP antibodies target the N-terminus of β-APP and congregate at regions of impaired axonal vesicle transport, which are associated with axonal injury. Zymed cites a recent study demonstrating that its C-terminus specific β-APP antibody has a higher signal-to-noise ratio than a β-APP anti-



Noise abatement: Invitrogen's new antibodies.

body specific for the N-terminus. It is suitable for use in ELISAs, western blotting, immunofluorescence and immunohistochemistry.

Troponin

Biodesign International www.biodesign.com

The heart of the matter

Troponin antibodies and antigens are available as I-Cardiac, I-Skeletal, T-Cardiac, T-Skeletal, C-Cardiac and Complex. Most of these antibodies have been characterized by their ability to recognize complex and free forms of the protein, and monoclonals and polyclonals have identified epitope specificities. The products are suitable for research use in immunoaffinity purification, Western blot and immunoassays. New monoclonal antibodies are available for immunochromatography. Recently introduced products include Mab to phosphorylated troponin I-Cardiac and Mab to dephospho-

rylated troponin I-Cardiac. A new myoglobin clone, creatine kinase, FABP, C-reactive protein and GPBB are also available.

Hand Saver

Stoelting

www.stoeltingco.com

Mix master

The Hand Saver pump automates the task of mixing adjuvant/antigen emulsions by forcing the fluid back and forth through a double-hub needle. Such microemulsions are needed for immunology work and involve manually pushing syringe plungers back and forth until emulsions are fully mixed and suspended. This can lead to repetitive strain injury. With the Hand Saver, syringes are mounted on the pump and left to mix automatically.

Antibody/ Assay brochure

Pierce

www.piercenet.com

Any questions?

Pierce Chemical is offering a free 40-page brochure that answers frequently asked questions related to antibody production, purification, modification and assay development. The brochure is set out in a question-and-answer format and includes tables of product comparisons. Go to the company's website for a copy, or visit www.antibodyquestions.com.

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

CUSTOMISED
immunoassays for
research and production



BioGenes

Gesellschaft für Biopolymere mbH

Internet www.biogenes.de

Berlin • Germany

Phone +49 (0)30-65 76 23 96

If your work requires reagents for:

~ proteasome ~
~ ubiquitin ~
~ COP9-signalosome ~
research

we have the most comprehensive range of
antibodies ~ biochemicals ~ kits

www.proteasome.com

AFFINITY Research Products Ltd.
Mamhead Castle, Exeter, EX6 8HD, U.K.
T. +44(0) 1626 891010 ~ F. +44(0) 1626 891090
~ Email info@proteasome.com ~

ADVERTISEMENT

**DIVERSE
SMALL
MOLECULES
READY FOR
SCREENING**



**10K-50K
Compounds
in Plates**

• Next Day Delivery*

• Competitively Priced

(*Some Restrictions Apply)

Chembridge Corporation • 800.964.6143

858.451.7400 • 858.451.7401 FAX • E-mail: chem@chembridge.com

www.chembridge.com

**Silver Stain Kit II
for Electrophoresis**

Rapid, sensitive detection
of proteins & nucleic acids

Wako

Wako BioProducts

Wako Chemicals USA, Inc.

1600 Bellwood Road

Richmond, VA 23237

Telephone: (804) 271-7677

Facsimile: (804) 271-7791

(800) 992-9256

bioproducts@wakousa.com

**Polyclonal Antibodies
& Custom Peptides**

Please contact us about complete
services for your custom peptide
and polyclonal antibody needs.

ResGen™

Invitrogen Corporation

U.S. or Canada 800-533-4363

FAX 256-536-9016

Check out our web site at
www.resgen.com

Contacts

Publisher: Fabien Savenay

Editor: Paul Smaglik

Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: classified@nature.com

Group European Manager:

Leonie Welss (4954)

European Manager:

Nevin Bayoumi (4978)

European Sales Consultant:

Jonathan Bull (4973)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andy Douglas (4975)

Holland/ Italy: Nevin Bayoumi (4978)

Scandinavia: Silje Opstrup (4994)

Production Manager:

Billie Franklin
To send films and materials use
London address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: classified@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau
Tel + 33 (0) 1 43 20 16 51
Fax + 33 (0) 1 43 20 51 52
e-mail: c.nioxchateau@nature.com

Spain/ Portugal:

Amanda Wren
Tel + 34 91 391 3835
Fax + 34 91 319 0091
e-mail: a.wren@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mails: p.phelan@nature.com
k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: classified@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Corissa Salzman

Japan Head Office, Tokyo

Shin-Mitsuke Building (4F),
3-6 Ichigaya Tamachi Shinjuku-Ku,
Tokyo 162
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.cowan@naturejp.com

Japan Manager: Kate Cowan

Immunological planets align

Sometimes a discipline's planets slide into perfect alignment and, at the moment, immunology seems to be such a case. The field offers scientific opportunity, meets human needs and is well funded, from both public and private sources.

Scientific opportunity comes in the form of the draft human genome sequence, combined with sequences of pathogens causing malaria and tuberculosis. But it is not limited to those applying genetics to infectious diseases. The emergence of 'discovery science' — such as collecting genomic data — only underscores the need for more 'hypotheses science' — conducting experiments both *in vitro* and *in vivo* (see *Nature Immunology* **2**, 373; 2001).

Human need for HIV/AIDS vaccines and drugs jump-started the field by triggering more investment from government and attracting interest from well-funded sources such as the Bill and Melinda Gates Foundation. But the World Health Organization and others have steadily made the case that diseases such as malaria and tuberculosis also warrant attention.

And it seems that some of the most exciting experimental medicine will rely on basic immunology to succeed. Islet-cell transplantation to combat diabetes has shown promise in a small group of patients. But it would be an even bigger boon if advances in immunology meant fewer anti-rejection drugs for recipients. Similarly, gene therapy would be safer and more effective if immunologists could find ways to trick the body into allowing viruses carrying therapeutic genes to persist.

Foundations and governments seem to be recognizing these problems, with money and infrastructure — both of which mean more opportunities for immunologists.

Paul Smaglik

Naturejobs editor



Contents

CAREERS AND RECRUITMENT

Demographics promises young immunologists strong prospects in Europe **p4**

Foundations and philanthropies galvanize US drug research **p6**

WWW.NATURE.COM/ NATUREJOBS

Career centre
Information on the scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



Immunological assay of Europe

Young European immunologists should benefit from demographic shifts if they can wait long enough for the opportunities to materialize, says Helen Gavaghan.

Marita Troye-Blomberg, a malaria immunologist at the Wenner-Gren Institute of the University of Stockholm, has two words of advice for PhD students and postdocs discouraged about finding permanent posts as immunologists: “Hang on.” The coming decade will see a wave of retirements by senior academics in basic and clinical immunology in Europe. “So many of us born around the 1940s are due to retire, and many of those born around the 1950s have already left science or moved abroad because we were occupying the top positions,” Troye-Blomberg says. “Replacements will be needed, but it will be competitive.”

They are, after all, working in perhaps the most mature and well-understood field of biology, says Fritz Melchers, director for 20 years of the Basel Institute of Immunology (which closed last year), but one where there are still some very basic questions to answer and much applied research to do. That should make for an

“Hang on”, Marita Troye-Blomberg (right), of Stockholm University, advises discouraged immunologists seeking career advancement.



Professorial tips

“Be prepared to take some risks, travel, stay focused on your research goals, but build up a diverse experience around your subject,” says David Pritchard, professor of parasite immunology at Nottingham University. “Apply for fellowships for research-leave to freshen your knowledge of contemporary developments, and don’t forget the history of immunology, because that is where many basic questions and clues to their answer lie.”

Pritchard has put much of this advice into practice and has followed a career path that should encourage some open mindedness.

He started out studying marine zoology and took a course in immunology in his final year, then a masters degree in immunology at Birmingham University. He then moved to a PhD in the immunology of parasites, and then to industry. He decided to reverse the usual career progression and to leave his permanent job for the vagaries of a postdoc. He won a lectureship after two years and, while climbing the career ladder, he took research fellowship leave. He now teaches basic immunology, and applies new tools to the research he first looked at 20 years ago when in industry.

H.G.

intriguing mix for career immunologists — the challenge of learning the huge body of existing work while prying loose answers in unknown territory (see “Professorial tips”, above).

INNATE QUESTIONS

“How does tolerance develop, how do regulator cells work, how do lymphocytes move in the body, what is innate immunity?” asks Martin Röhlhoff, from the Institute for Microbiology and Immunology in Erlangen, Germany. Such questions are of long standing within immunology, and because answers do not always have obvious applications it can sometimes be difficult to get funding, says Troye-Blomberg. But

the answers will eventually serve a wide range of medical need. Immunology encompasses and underpins a breathtaking array of medical topics — allergic reactions, autoimmunity, rheumatology, vaccinology, oncology, transplantation, infectious diseases and immune deficiencies.

“There is much more of an interface between basic research and clinical immunology in a number of fields than in any other therapeutic area,” says Jean-François Bach, an immunologist at Necker Hospital, Paris.

So how will immunologists solve their basic questions as well as bridging the gap between basic and clinical research? The tools of molecular biology — proteomics, bioinformatics, transgenic mice and the like — will, of course, help, and should be pressed into service, says Bach. But he cautions against placing them centre stage. Repeatedly, he and other immunologists say that it is time now to think of immunology in terms of the whole system, not its parts.

Having spent 10 to 20 years deconstructing the immune system all the way to DNA and RNA, the time has come to reconstruct it, says Geneviève Milon from the Institut Pasteur in Paris and chair of the education committee of the International Union of Immunological Societies.

MOVING *IN VIVO*

The only way to explore fully the mysteries of the immune system as well as understanding its interactions with microorganisms is to study immunology *in vivo*, says Marie-Christine Béné, an immunologist at Nancy University, France.

Others around Europe echo the view. “Laboratory models of infection have some value but sometimes study the immune systems of inappropriate hosts. You need to go into the field, collect samples from people, survey the village and assess the parasite burden in parallel with work in models,” says David Pritchard, professor of parasite immunology at Nottingham University. Such a holistic approach is the only way to learn something of the parasite’s interaction with its host in the real world.

Indeed, field work combined with sequence data, bioinformatics and the panoply of molecular biology’s tools are a powerful combination. Immunology therefore presents an opportunity for the synthesis of many new and old biological techniques and has the ambitious aim of explaining a physiological system that pervades the body and, when not functioning normally or when under attack, underpins many diseases.

SUPPORTING STRUCTURE

So how well does the scientific infrastructure across Europe support the challenge? On the whole, quite well, although closure by Roche of the Basel Institute for Immunology was a severe blow to the community. For 30 years, the institute was an intellectual powerhouse of thinking about the basic questions of immunology and it was closed just as the need to think once again holistically was being acknowledged.

And there are some scattered clouds on the horizon, one being the availability of professors of immunology. France, for example, is well stocked with clinical immunologists, but not professors of basic immunology (see “Clinical immunology paths”, right). “This is a

bizarre development,” says Béné, and one that Röllinghof says he could see developing also in Germany.

It is doubtful, though, that these clouds will build into a storm front, and they could force some interesting international mixing, because, on the plus side, there will soon be room at the top, a practical consideration for any career.

There is, of course, the usual problem that academics are paid less than people in industry or finance, but when asked if she had anything to say to aspiring immunologists, Béné said simply, “Tell them it’s great.” ■

Helen Gavaghan is a freelance journalist based in West Yorkshire.



Immunologist Martin Röllinghof wrestles with some of the long-standing and fundamental questions in immunology.

Web links

The European Federation of Immunological Societies (links to all national societies in Europe) ♦ <http://www.efis.org>

International Union of Immunological Societies

♦ <http://www.qimr.edu.au/iuis>

Institut Curie ♦ <http://www.curie.fr>

Institut Pasteur ♦ <http://www.pasteur.fr>

The British Society for Immunology (including job vacancies)

♦ <http://immunology.org>

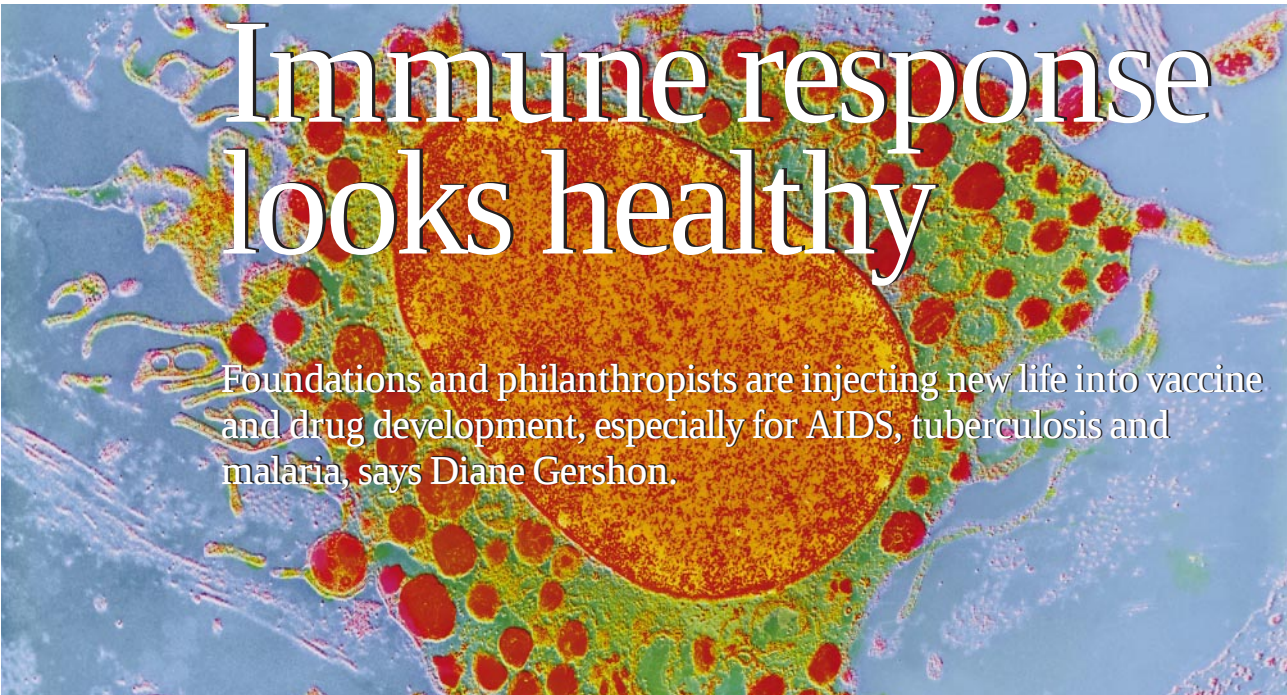
Clinical immunology paths

Clinical immunology divides into two branches: laboratory medicine and the bedside. Within laboratory medicine, clinical immunology has a well-defined career path. Clinical immunology at the bedside is less well defined and is not a medical specialty, except in Britain, the Netherlands and France.

There are clinical immunologists elsewhere in Europe, but there is no agreed training path. “In some countries,” says Marie-Christine Béné, from Nancy University, France, and co-ordinator of the clinical immunology group at the European Federation for Immunological

Societies, “immunology is recognized as a specialty, in others it is part of internal medicine.”

Béné argues that recognizing the discipline as a distinct medical specialty will give it higher visibility and help career mobility in Europe. Fritz Melchers, president of the International Union of Immunological Societies, thinks the hard fight for acceptance is necessary for development of the discipline. For example, autoimmunity, like cancer, affects all tissue types. To understand its basis, he says, immunologists need to work with patients with all types of autoimmune disease. H.G.



Immune response looks healthy

Foundations and philanthropists are injecting new life into vaccine and drug development, especially for AIDS, tuberculosis and malaria, says Diane Gershon.

The health-research landscape — and with it, careers in immunology — is changing, thanks in part to hundreds of millions of dollars pledged by foundations and philanthropists to combat infectious diseases including AIDS, tuberculosis (TB) and malaria.

Many foundations are increasing donations to help fight infectious diseases, but in recent years, the Bill and Melinda Gates Foundation, which has an endowment of \$23.5 billion, has emerged as the leader. The Gates Foundation originally focused on using existing vaccines in developing nations. But some of its newer initiatives “have new vaccines as their goal”, says Gordon Perkin, director of the global health programme at the Gates Foundation.

The increase of foundation support for vaccine development could work synergistically with

Foundations pressure public research

The activities of philanthropic foundations have lit a fire under publicly funded research into infectious diseases, says Tony Fauci, director of the US National Institute of Allergy and Infectious Diseases (NIAID). This has led to increased urgency in the basic science that enables new approaches to be taken to vaccines, he says.

Most foundation money tends to be dedicated to vaccine development. “That, to me, gives you even more incentives to turn the afterburner on in basic research,” Fauci says.

In recent years, the NIAID has grown quickly within the US

National Institutes of Health (NIH). It was once the eighth-largest NIH institute by budget — it now ranks third. Some of that growth can be attributed to the infusion of federal money in HIV/AIDS research, but Fauci notes that AIDS-related and non-AIDS-related research funding has grown at roughly the same rate since 1990.

The increase in funding has meant more support for both intramural and extramural research. This year the NIAID has received \$531.6 million for research — almost 90% of which goes to extramural grants, with the remainder spent on

research at the NIH campus. Of that figure, less than a third is spent on clinical research. The ratio between basic and clinical research at the NIAID underscores the way publicly funded institutes complement philanthropic foundations — the government supports fundamental science, whereas foundations drive applied research.

The increase in both kinds of work is good news for young scientists considering a career in immunology. “I can say without hyperbole that opportunities now in immunology are really unprecedented,” Fauci says.

Paul Smaglik

increases in the budget from the US National Institutes of Allergy and Infectious Diseases, which has seen its research funding grow to \$531 million in 2001.

National Institutes of Health (NIH) grants tend to fund more basic work (see “Foundations pressure public research”, above), whereas some of the more

recent foundation-funded initiatives are more applied. The increase in both kinds of support bodes well for young researchers looking to enter a vibrant field — especially in diseases that are prevalent in developing countries but which have largely been ignored in Western ones (see “Foundation helps fund hookworm enthusiast”, overleaf).

A SENSE OF OPTIMISM

“The prospect of significant and sustained funding has begun to attract bright young scientists, naturally concerned about their own long-term careers, to devote their time and attention to these issues,” says Alfred Sommer, dean of the Bloomberg School of Public Health at Johns Hopkins University in Baltimore, Maryland. “In the long term that may prove as important as the research these grants are funding today,” he says.

Mary Galinski, of the division of infectious diseases in Emory University’s Department of Medicine, and an affiliate scientist at the nearby Yerkes Regional Primate Research Center, can attest to how a foundation’s interest can help. When Galinski founded the Malaria Foundation International about a decade ago to help coordinate research and galvanize support, little attention was being paid to malaria.

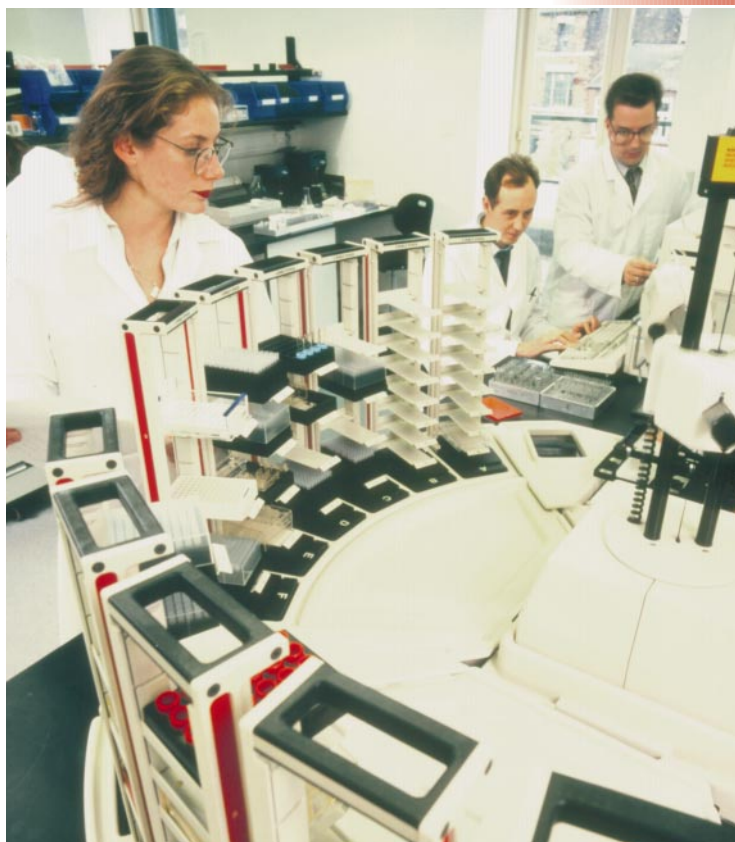
Funding — if anything — was decreasing, she says. “We saw people leaving the field or splitting their programmes between malaria and HIV,” says Galinski, of her days as a junior faculty member at the New York University Medical Center. Galinski moved to Emory 2½ years ago to build a malaria programme at the university’s Vaccine Research Center in Atlanta, Georgia. That programme was bolstered when the vaccine centre landed an umbrella contract with the Malaria Vaccine Initiative (MVI), which the Gates Foundation launched in 1999 with a \$50-million grant. Galinski will head efforts to test malaria candidate vaccines for MVI in non-human primates.

There is now a “sense of optimism”, Galinski says. But she adds that current funding levels for malaria vaccine research are “still relatively low-end”, compared to HIV, although the research challenge is no less formidable. Galinski also believes that a re-evaluation of MVI’s current list of candidate vaccines may be necessary in light of the malaria genome project.

Another philanthropist-funded project will do just that. The new Johns Hopkins Malaria Institute, which was announced in May after an anonymous donor gave \$100 million towards its creation, will be set up specifically to embrace new scientific tools, such as genetic sequencing and bioinformatics. The intent is to take a fresh look at the malaria problem, with the goal of developing innovative vaccines and new antimalarial drugs.

It will also provide more jobs. Twelve new investigators will be recruited within the next five years; new hirings will not be limited to malaria experts. Those new recruits, along with the four or five groups that already exist on campus, will make up the core of the new malaria institute, says Sommer. The institute will also benefit from its close proximity to NIH, which has been a long-time supporter of malaria research.

Sommer notes that in addition to funding high-risk approaches that the NIH tends to shy away from, these awards provide a level of visibility to these neglected



CROWN COPYRIGHT/HEALTH & SAFETY LABORATORY/SPL

areas that is drawing others to the table. The day the malaria institute was announced, Sommer was contacted by a well-known philanthropist who wished to know who the anonymous donor was so that he could call and “personally thank him”. A week later, he was called by another well-known philanthropist who asked: “What is all this buzz about malaria and tuberculosis?”

The prospect of significant and sustained funding has begun to attract more bright young scientists.

BUZZ ATTRACTS FUNDING

Whatever the source of the buzz, it is certainly bringing more attention — and money — to infectious-disease research. Tuberculosis research is also picking up, thanks to foundation attention, says Gilla Kaplan of the Laboratory of Cellular Physiology at Rockefeller University, New York. Kaplan is one of a dozen or so core scientists funded through the Sequella Global Tuberculosis Foundation. The foundation was established in 1997 to accelerate the clinical testing of potential TB vaccines and is funded in part by the Gates Foundation. The Sequella initiative “has crystallized and focused the effort towards vaccine development in a much more practical and pragmatic way”, says Kaplan.

Scientific progress is also fuelling foundation interest and funding in TB drug development. The recent sequencing of the *Mycobacterium tuberculosis* genome, and limitations in current TB treatment strategies led to the creation last year of the Global Alliance for TB Drug Development. The alliance operates as a small, virtual R&D company and already has financial commitments totalling \$40 million from the Gates and Rockefeller foundations. Its mission will be to accelerate the development of TB drugs that will shorten the duration

of treatment, improve care of latent TB infection and be effective against multi-drug-resistant TB.

Although still in its early days, the alliance will focus on “picking up compounds that have come through some of that early development process and that might be ready for phase I clinical trials”, says Rick O’Brien, chief of research in the division of TB elimination at the Centers for Disease Control in Atlanta, and chair of the alliance’s scientific advisory committee. The goal is to have a major new TB drug licensed by 2010.

Andrew McMichael and his team are also learning how to function more like a pharmaceutical company. McMichael, from the MRC Immunology Unit at Oxford University, heads one of the five groups with funding from the International AIDS Vaccine Initiative (IAVI). His group’s experimental HIV-1 vaccine is currently being safety tested in Britain and Kenya. Taking a laboratory vaccine into clinical trials “wouldn’t have been possible without the IAVI’s support”, he says. Just to get the phase I trials up and running McMichael has had to employ eight more people in Oxford and about the same number in Kenya as project managers, research nurses, research physicians, technicians, data-entry clerks and AIDS counsellors.

Although the influx of new funds from philanthropists is considerable and has brought increased visibility to previously neglected areas of research, the initiatives they are funding are still relatively small scale. Seth Berkley, president and chief executive of the IAVI, notes that its budget is less than

1% of the NIH AIDS budget. “But we are the second-largest AIDS vaccine effort in the world, which is shocking if you think about it,” he says. The war on these diseases is far from over but the battle is in its early stages, says Berkley, in terms of getting people to take it seriously.

Diane Gershon is Assistant Editor, *Nature Medicine*, New Technology.

Web links

The Bill & Melinda Gates Foundation

♦ <http://www.gatesfoundation.org>

The Rockefeller Foundation ♦ <http://www.rockfound.org>

Alfred P. Sloan Foundation ♦ <http://www.sloan.org>

The International AIDS Vaccine Initiative ♦ <http://www.iavi.org>

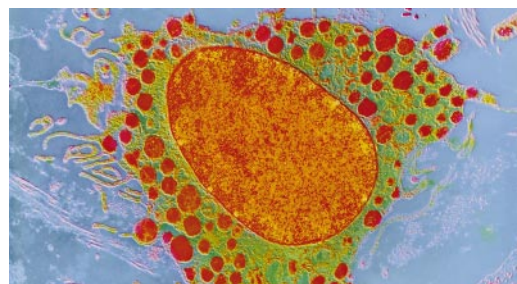
Malaria Vaccine Initiative ♦ <http://www.malariaivaccine.org>

Global Alliance for TB Drug Development ♦

<http://www.tballiance.org>

Sequella Global Tuberculosis Foundation

♦ <http://www.sequellafoundation.org>



S.CINTI/CNRI/SPF

Foundation helps fund hookworm enthusiast

Even at the age of 13, Peter Hotez — who kept a copy of *Manson's Tropical Diseases* by his bed — was hell-bent on becoming a parasitologist. He wanted to work on a parasitic disease that was a major problem in developing countries, yet one that was under-explored scientifically. While in the MD/PhD programme at Rockefeller University/Cornell University Medical College, he picked up a copy of a lecture given in 1962 on hookworm, *The Great Neglected Disease of Humankind*, did a literature search and discovered there was scant information on the organism. He was hooked.

Hookworm (*Ancylostoma*) affects about 1 billion people worldwide — almost 200 million in China alone, where Hotez has a collaboration with the Institute of Parasitic Diseases in Shanghai. Although usually not fatal, these blood-feeding parasites infect people

of all ages, causing individuals to suffer chronic blood loss. Infections are particularly debilitating in children where they lead to stunted growth and intellectual retardation.

From 1989 until last August, Hotez ran a basic science lab on hookworm research at Yale University. “I struggled at Yale,” he says. “I always had continuous funding but there were always enormous hoops I’d have to jump through.” Hotez has received some support from the US National Institutes of Health, as well as private foundations such as the March of Dimes Birth Defects Foundation, the American Heart Association and the Culpepper Foundation — not usually known for their support of parasitology. But his big break came last year with an \$18-million award from the Bill and Melinda Gates Foundation, which will provide the kind of boost necessary to propel his research to the next

stage — the clinical testing of a hookworm vaccine.

Although the grant was issued to the Albert B. Sabin Vaccine Institute, Hotez is the principal investigator and initial research is being carried out in his lab. Hotez moved from Yale to the George Washington University Medical Center in Washington, DC, last summer to become the newly appointed chair of the university’s Department of Microbiology and Tropical Medicine, which he is restructuring and expanding. The move also allowed Hotez to be closer to his adviser, Phil Russell, at the Sabin Institute’s satellite office in Rockville, Maryland. He is also closer to organizations with interests in foreign policy and to the Walter Reed Army Institute of Research, which has experience in pilot vaccine production.

In order to make a vaccine, you have to do some very

pedestrian things, says Hotez, such as re-engineering the same protein in different expression vectors and making the protein soluble. Ordinarily, this could be done by industry, but hookworm is a disease “that’s below the radar screen for almost any drug company”. Hotez did make the rounds of pharmaceutical companies, but remarks: “Somebody once said to me, the key to success in business is to know how to distinguish between a cup of coffee and a certified cheque.” Companies were empathetic, but he says he got lots of cups of coffee and not much else. “That’s the blessing of the Gates Foundation ... it treads where no one else will.” Hotez has identified a class of antigens from the infective larval stages of the parasite that offer promise in eliciting vaccine immunity and expects to begin clinical testing of a candidate vaccine within two years.

D.G.